Texas Medical Center Library
DigitalCommons@TMC

UT GSBS Dissertations and Theses (Open Access) | Graduate School of Biomedical Sciences

4-2018

Discovery and effects of pharmacological inhibition of the E3 ligase Skp2 by small molecule protein-protein interaction disruptors

John K. Morrow

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the Medicinal Chemistry and Pharmaceutics Commons, Pharmaceutics and Drug Design Commons, Pharmacology Commons, Structural Biology Commons, and the Therapeutics Commons

Recommended Citation
Morrow, John K., "Discovery and effects of pharmacological inhibition of the E3 ligase Skp2 by small molecule protein-protein interaction disruptors" (2018). *UT GSBS Dissertations and Theses (Open Access)*. 861.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/861

This Dissertation (PhD) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@TMC. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact laurel.sanders@library.tmc.edu.

Discovery and effects of pharmacological inhibition of the E3 ligase Skp2 by small molecule protein-protein interaction disruptors

by

John Kenneth Morrow, M.S.

APPROVED:

_____
Advisory Professor, Shuxing Zhang, Ph.D.

_____
Varsha Gandhi, Ph.D.

_____
Scott Gilbertson, Ph.D.

_____
Zhimin Lu, Ph.D.

_____
Giulio Draetta, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

Discovery and effects of pharmacological inhibition of the E3 ligase Skp2 by small molecule protein-protein interaction disruptors

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

In Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

John Kenneth Morrow, M.S.

Houston, Texas

May 2018

Dedication

This work is dedicated to my mother, Jeanne Morrow, who has always been supportive in my journey through my scientific career, and love for me knows no bounds. My father Philip Ross Morrow, whose passion for science and sudden passing from a rare form of cancer is my primary motivation to go into this field. Finally, to Angie Marie Torres-Adorno, who has loved, inspired, encouraged, motivated and supported me on whatever decision I have made.

Acknowledgements

I would like to acknowledge all the hardworking researchers in Dr. Shuxing Zhang's lab, Dr. Shuxing Zhang himself who has mentored me through the last nine years, and who pushed me to achieve great things. Also other lab members including: Rajan Chaudhari, Beibei Huang, Michael Cato, Zhi Tan, Lu Chen, Srinivas Reddy Alla, Hoang Tran, Lei Du-Cuny, Longzhang Tian, Sharangdhar Phatak, and Nathan Ihle. I owe great acknowledgement to my advisory committee, for their valuable inputs and insights into this thesis. I could not have done it without your support. I also thank all the members in Hui-Kuan Lin's lab, especially Lori Chan, and the members of Dr. Naoto Ueno's lab; all are great researchers for their dedicated contributions to the collaborative efforts in their experimental validations.

I also acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources that have contributed to the research results. URL: http://www.tacc.utexas.edu

Discovery and effects of pharmacological inhibition of the E3 ligase Skp2 by small molecule protein-protein interaction disruptors

By John Kenneth Morrow, M.S.

Advisor: Shuxing Zhang, Ph.D.

Skp2 (S-phase kinase-associated protein 2), one component of the SCF E3 ubiquitin ligase complex, directly interacts with Skp1 and indirectly associates with Cullin1 and Rbx1 to bridge the E2 conjugating enzyme with its protein substrate to execute its E3 ligase activity. Skp2 is an Fbox protein (due to it containing an Fbox domain) and it is the rate-limiting component of the SCF complex. Skp2 targets several cell-cycle regulatory proteins for ubiquitination and degradation; most notable and significant for cancer are the cyclin-dependent kinase inhibitor, p27. Skp2 is an oncogene and studies have shown that over-expression of Skp2 leads to increased degradation of p27 and increased proliferation in several tumor types. Additionally, Skp2 is over-expressed in multiple human cancers. Clearly, Skp2 represents an attractive target for attenuating p27 ubiquitination and subsequent cell cycle progression. However, Skp2 does not have an easily identifiable and druggable "pocket" on which small molecules can bind; it interacts with Skp1 through the Fbox domain and binds to an accessory protein called Cks1 to bind to p27. Despite this hurdle, in this study, two selective small molecule inhibitors of the Skp2 SCF complex were discovered via an *in silico* screen that disrupt two places: the Skp1/Skp2 interaction site and the p27 binding site via targeting hot-spot residues.

The Skp1/Skp2 inhibitor disruption resulted in restoring p27 levels in the nucleus and blocks cancer progression and cancer stem cell traits. Additionally, the inhibitors phenocopy the effects of genetic Skp2 deficiency. Two specific residues on Skp2 were predicted to bind to this Skp1/Skp2 inhibitor: Trp97 and Asp98. When these residues were mutated to alanine, the inhibitor lost its ability to bind to Skp2. To investigate the flexibility and understand the conformational change upon inhibitor binding and dynamics of the SCF complex, molecular dynamics simulations, homology models, and structural analysis was carried out on the complex with and without the inhibitors. These simulations showed that the contributions of the N-terminal tail region of Skp2 does not contribute directly to the binding of these inhibitors; but its conformation is important in the context of the other members of the SCF complex. Further dynamics analysis validated the mutagenesis results, showing that the two Skp2 mutants (Trp97Ala, Asp98Ala) that retained Skp1 binding but blocked inhibitor binding were stable, whereas the mutant that was unable to retain Skp1 binding (Trp127Ala) showed destabilization in the Fbox domain. Finally, active recruitment events after post-translational modifications are shown to be possible by the interaction of phosphorylated Ser256 on Skp2 with Lys104 loop region on Cul1 The model shows that this is due to the significant flexibility in the F-box domain of Skp2, making this interaction very likely. These results show that Skp2 is a promising target on which protein-protein interaction disruptors can be designed, and consideration of the dynamics of protein complexes is required to understand ligand binding.

Table of Contents

List of Illustrations

List of Tables

Abbreviations

PPI: Protein-protein interaction

Ub: Ubiquitin

GPCR: G-protein coupled receptors

PDB: Protein Data Bank

SCF: Skp1-Cullin-F-box

HTS: High-throughput Screening

ZL-25: ID of small molecule disruptor of Skp1/Skp2

ZM-276, ZM-405: ID of small molecule inhibitors of Skp2/Cks1/p27

MD: Molecular Dynamics

RMSD: Root mean square deviation

RMSE: Root mean square error

QSAR: Quantitative Structure Function Relationship

H-bond: Hydrogen bond

vdW: van der Waals

PCA: Principle component analysis

ANOVA: Analysis of variance

µL: microliter

µM: micromolar

Chapter 1: Introduction

1.1 Targeting Protein-Protein Interactions to Expand the Drug-Target Space

1.1.1 Current state of Small Molecule Cancer Therapies

Despite a reported spending of $50 billion on research in the pharmaceutical industry into small molecule therapeutics (reported in 2006), there is a surprising lack of information on the known universe on which these therapeutics act upon. It has been proposed in 2006 that the consensus number of drug targets for all classes of therapies that have been FDA approved numbers only 324 (1). This estimate is rather liberal when compared to other studies that use slightly different metrics and at a higher but still modest count of 667 (2). When one breaks down the approved drugs based on the target type, the overwhelming majority of them target very few major classes: G-protein coupled receptors (GPCRs), nuclear receptors, ion channels or enzymes (most prominent of the enzyme category are the kinases) (1, 3) (**Figure 1.** Distribution of Approved drugs sorted by target type. Enzymes and GPCRs make up over 2/3 of all targets.. More recent and comprehensive studies in the last year place

the human-genome derived practical target count pathway analysis that put the number at 273 (4), or even 218 (3).

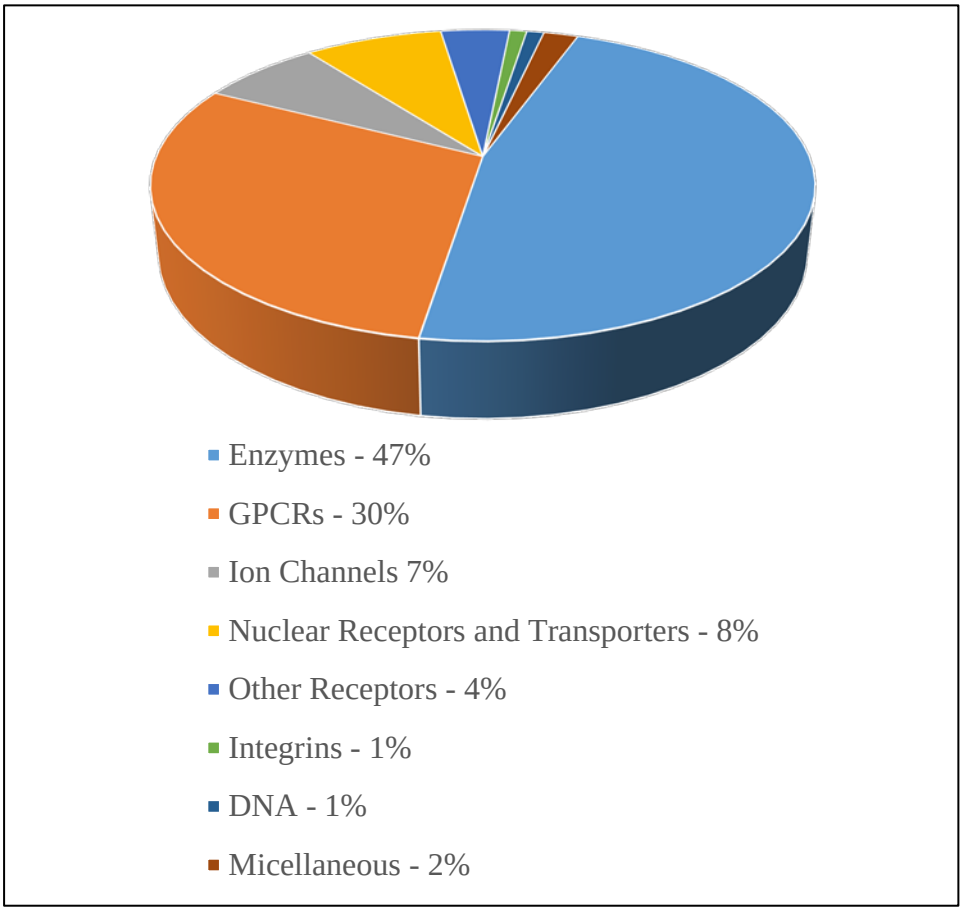


Figure 1. Distribution of Approved drugs sorted by target type. Enzymes and GPCRs make up over 2/3 of all targets.

The breakdown of the ChEMBL database conducted as of writing shows that out of 7,610 total targets, 3983 or 52% are in the “enzyme” category (5). This paradigm follows the classic drug-target model of considering a target druggable by targeting the active-site of the protein and blocking natural ligands of these receptors/enzymes. This becomes problematic when one considers that the non-enzyme classification of proteins expressed in the proteome make up the majority of proteins involved in signaling pathways, transport, transcription, translation and other

vital cellular functions. Clearly then the state of small molecules remains very narrowly focused on archetypes and models that can be argued as outdated especially when one considers that many diseases, most notably cancer, remain best described as a disease “at best, minimally controlled by modern medicine” (6).

1.1.2 PPIs and Their Potential as a Therapeutic Target

The inherent complexity of cancer (and other diseases that have multiple pathways involved in their phenotype and have poor prognoses) leads us to the conclusion that the paradigm of “one drug, for one target, for one disease” (also sometimes referred to as “low-hanging fruit” is not capable of reducing the disease burden. One must expand the drug-target space and consider unconventional target types.

To expand this drug-target space and take into account the functions of proteins in their relevant pathways, it is critical to consider that most cellular proteins do not act as isolated units; they often form complexes with other proteins (7). Formation of these protein complexes require specific interactions that become the foundation for almost all biological processes, especially signal transduction. The complexity of these protein-protein interactions (PPIs) is what primarily provides them with their diversity in functions. From this complexity, distinctions can be made in the shape, size and surface complementarity of protein-protein interactions (8-12). These interactions can be described as locally optimized, with the clustered, networked, densely packed residues contributing mainly and cooperatively to the stability of the complex (13). The

sites by which proteins interact with their partners are formed by surfaces with adequate electrostatic complementarity and shape (14-17). Other major factors that influence protein-protein interactions are hydrophobicity (9, 18) and electrostatic interactions (16), and flexibility (14-16). While interfaces with an area as small as 1150-1200 $Å^2$ have been described in low-stability and short-lived complexes, "standard-size" interfaces (19) are roughly 1600 $Å^2$ (+/- 400 $Å^2$) and involve only small changes in conformation (12) and approximately 750-1500 $Å^2$ of the surface area buried in each protein (20). The "large" interfaces bury 2000-4660 $Å^2$ and occur mostly between G-protein and other components of signal transducers, and between proteases and one class of their inhibitors (17, 19). Understanding the connection between structure and function of molecular systems can be achieved through examination of the protein-protein associations, and enables the characterization of their energetic properties (21, 22). When one considers that there are over 45,000 protein-protein pairs in yeasts, it should become apparent that the number of potential, druggable targets in humans is enormous (23).

Cellular dysfunctions from faulty protein-protein interactions are the underlying cause of a variety of diseases, including cancer and neurological disorders (24, 25). Despite the critical importance and relative abundance of protein-protein interactions, very few small-molecule PPI inhibitors have made the difficult transition from hits to marketed drugs. While peptide-based inhibitors derived from 5-50 amino acids of a single member of an interacting protein pair are easy to construct, conversion of these peptides to "drug-like" molecules has proved to be the biggest hurdle (26, 27). One

significant part of this problem is the nature of the PPIs; classic protein inhibitors primarily target well-defined grooves and pockets (28, 29), whereas PPIs are usually large and relatively planar surfaces that were originally considered not conducive to druggability by "drug-like" molecules (16, 30). Advancements in two areas: the understanding of the physical nature of PPIs as well as the increasing diversity and quality of small, lead-like compounds have allowed PPIs into the realm of valid targets, thereby greatly expanding the available target space for a plethora of diseases (31, 32).

1.2 Hot spots as a Basis for Drug Discovery Efforts

Used with explicit permission from Bentham Science Publishers in the article: Current Pharmaceutical Design. 2012;18(9):1255-65.

1.2.1 Introduction to hot spots

It has been well established that for all proteins, the energy distribution is not uniform across a given protein-protein interaction; that a small subset of residues will have a more significant contribution to the binding free energy than other residues (13, 33-40). Clackson and Wells' pioneering study of the binding of human growth hormone to its receptor discovered these important residues and coined the term 'hot spots' (33). Later work revealed that hot spots occupy only a fraction of the larger interface area (34), and proclaimed a more precise definition of a hot spot as a residue whose mutation to alanine results in a decrease of at least 2.0 kcal/mol in binding free energy ($\Delta\Delta G_{binding}$) (41). Binding free energy, or $\Delta\Delta G$, is defined as $\Delta G^{mut} - \Delta G^{wt}$, where ΔG^{wt}

and ΔG^{mut} are the binding free energies upon complex formation of the wild-type and alanine-mutated proteins, respectively. It has been calculated that 9.5% of interfacial residues are hot spots (42).

The composition of hot spots is distinctive and not random, with tryptophan (21%), arginine (13.3%), and tyrosine (12.3%) being the only three fundamental amino acids having more than 10% frequency (34, 43). The complex formed by human growth hormone and its receptor is shown with its dominating tryptophan residues (**Figure 2**). Out of the 29 interfacial residues in this complex, only four hot spots have $\Delta\Delta G_{binding}$ greater than 4.5kcal/mol, and two of them are tryptophan. Tryptophan's unique function can be partially explained by its large and aromatic π-interactive nature (44) as well as its large hydrophobic surface and protective effects from water (45). When the bulky tryptophan residue is mutated to alanine, the difference in size generates a large cavity that creates a highly complex destabilization (34).

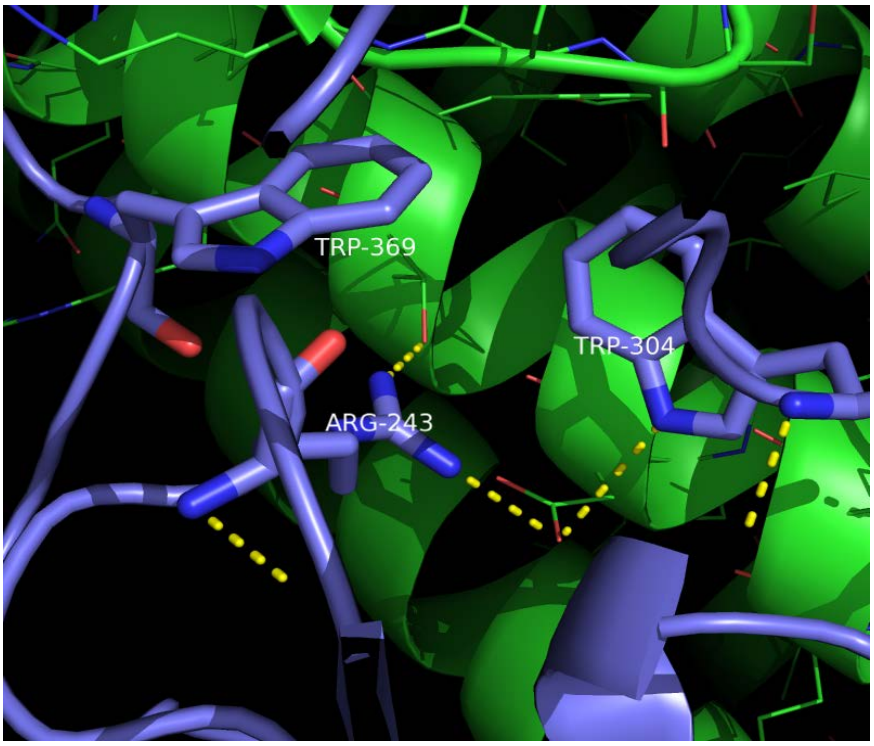


Figure 2. Example of typical hot spot residues of human growth hormone (green cartoon) and its receptor (blue sticks) (PDB Code 1a22). Three of the four hot spot residues of the receptor are represented in sticks, hydrogen bonds are shown in dashes. This result was displayed using PyMOL (46).

Consideration of the structural aspects of hot spots (47, 48) reveals that they are cooperative and structurally conserved (13, 36, 49, 50). This structural conservation is also apparent when examining the rate of mutation of surface residues; protein interfaces (51) and functional sites (52) mutate at a relatively slower rate when compared to other surface residues. Mutations of the interactions between proteins may occur largely by coevolution, where substitutions in one protein trigger reciprocal changes in the other protein (51). Numerous studies of protein-protein interfaces have revealed the presence of hot spots where binding affinity and specificity can be resolved by an epitope that is consisting of only a small portion of interfacial residues (34, 35, 53-55). Clearly, the structural conservation of hot spots and their correlating binding affinity make them attractive drug targets for small molecule inhibition. It is not

surprising then, that hot spots have been considered in the design of small molecule inhibitors of unwanted protein-protein interactions, and several potential drugs show great promise in targeting hot spots (20, 25, 42, 56-63).

1.2.2 Hot spots as potential drug targets

Hot spots can facilitate drug design in two ways. First, the binding site can be predicted by the presence of hot spots, and this binding site can be considered a starting point to perform docking and/or screening of ligands (64). Second, rigid docking could exploit the relatively less flexible hot spots, and improvements in protein docking would be achieved by choosing dominant conformation of the hot spot side chains resulting from molecular dynamics (MD) simulations rather than the unbound X-ray conformation (65, 66). Initial studies have shown that hot spots have physico-chemical properties that may be predictive (42, 53), and this prediction goes a long way to decipher the functional mechanism of proteins. From the mechanisms, valuable information can be used for a wide range of purposes: mutant proteins can be designed that will validate the interactions, drugs can be developed that target the protein-protein interactions, and molecular recognition can be expanded to help us understand complex signaling pathways.

Successful identification of hot spots involved in protein interfaces is crucial in determining potential active sites, and the corresponding druggability of a protein target. It is only after the completion of this step that rational, structural based drug design can begin. Systematic mutagenesis of protein-protein interactions has yielded a

wealth of information on the hot spots themselves, but there are significant inroads that can be made as to the general nature of these hot spots. Unfortunately, no general patterns of shape, charge or hydrophobicity exist that can be used to predict which residues form hot spots (12, 16). Additionally, residue conservation is rarely sufficient for accurate prediction of protein interfaces (50). Further potential pitfalls become apparent when considering the inherent plasticity and expansive diversity of PPIs.

Despite these challenges, the increasing use of computational methods holds high potential to provide accurate analysis in prediction of hot spots at a fraction of the time and cost for conventional, experimental techniques. Advancement in both hardware processing power as well as software predictive models will increase the speed and the accuracy of hot spot predictions. The increasing rate of both crystallographically resolved protein-protein complexes as well as experimental alanine mutagenesis studies will provide a more comprehensive dataset from which to further increase the accuracy of computational prediction of hot spots. Composite approaches that incorporate initial computational predictions, and are then validated against experimental means provide the most comprehensive understanding of the nature of hot spots. These validations will vastly improve the predictive power of *in silico* techniques by continually evolving from the increasing pool of data from which they are built upon.

1.2.3 Current Experimental Methods to predict hot spots

Experimental identification of hot spot residues is primarily performed by alanine scanning. This process involves mutation of a residue of interest to alanine, and

recording the resulting binding energy changes. If this mutation results in a marked drop in the binding constant (typically tenfold or higher), the residue is considered a hot spot (53). Substitution with alanine removes all atoms in the side chain past the β-carbon. Alanine is used because of its relatively inert methyl functional group without contributing additional flexibility (67-69). Mutation to glycine would also remove the side chain, but is not used since it can introduce unwanted conformational flexibility in the protein backbone (70).

In the context of protein folding, a mutation on a hot spot residue can be considered to destabilize the bound ensemble state relative to the unbound one (71, 72). It follows then that alanine scanning determines the specific contribution of residues to the ensemble's stability, and resulting protein function.

If we are to consider the plasticity of protein interfaces, alanine scanning mutations affect free energy surfaces that can lead to differential effects of the bound, unbound state, or both (**Figure 3**). It becomes advisable then to interpret both the structure and energetic properties of each ensemble to rather than a specific residue interactions (53). Initially, alanine scanning was applied towards human growth hormone and the growth hormone binding protein (33), and this technique has proved invaluable in PPI analysis and hot spot detection (12, 34, 39, 53, 61). The resulting data from alanine scanning can be deposited in the Alanine Scanning Energetics Database (ASEdb), and verified experimental hot spots from the literature are found in the Binding Interface Database (BID) (41, 73). These repositories, while useful, have two drawbacks. First, hot spot information from experimental studies is limited to very few

complexes. Secondly, it is not recommended to interpret this data to specific residue interactions (53).

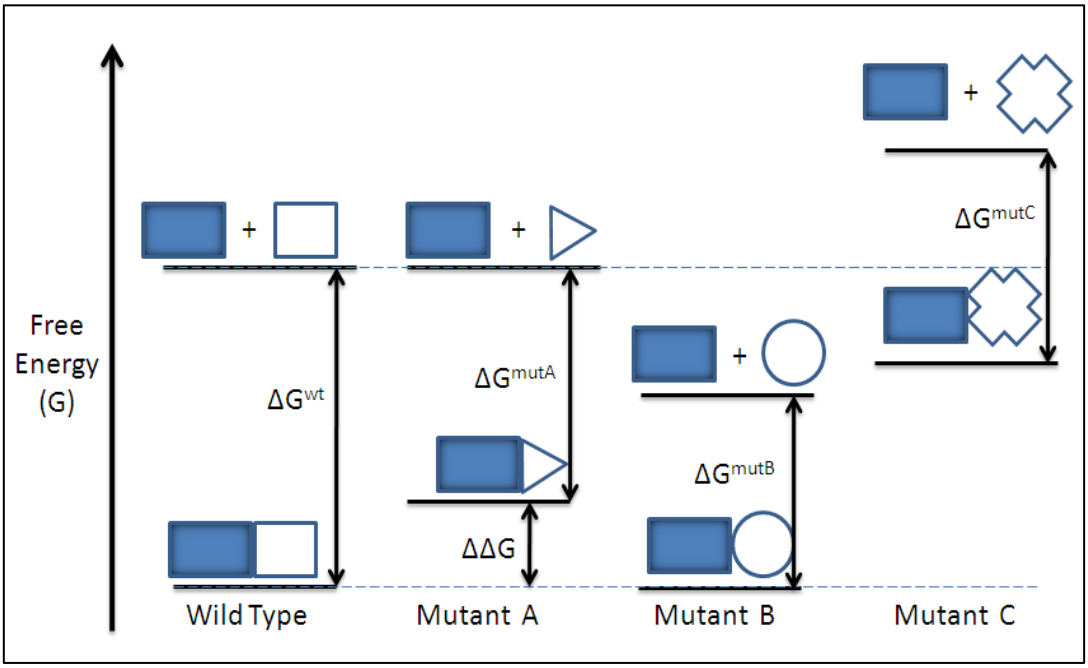


Figure 3. Mutations from alanine scanning can affect free energies of bound and unbound states. Mutant A is an ideal result where only the complex is destabilized, and the binding free energy with respect to the wildtype (ΔΔG) can be attributed to specific changes in the PPI contacts. Mutant C affects both states simultaneously. The measured binding free energy difference from B and C mutants must be interpreted cautiously. (Figure adapted from (53): "Unraveling hot spots in binding interfaces: progress and challenges", used with explicit permission from Current Opinion in Structural Biology).

Experimental mutagenesis of target proteins for elucidation of hot spots is not applicable on a large scale since individual mutants must be purified and analyzed separately (53). Systematic analysis requires each alanine-mutated protein to be constructed, placed in an expression system sometimes refolded, then the resulting protein activity is assessed in an *in vitro* assay (70). While techniques such as reflectometric interference spectroscopy (74) and "shotgun scanning" (75) have alleviated some of the labor involved, experimental hot spot analysis is still very time consuming and expensive. The theoretical and computational prediction of hot spots has become one of the most attractive and challenging topics in biochemistry,

biophysics, and bioinformatics, and each of these disciplines offer unique strategies to confront this challenge. Next, we will outline the methods and tools available for computational hot spot prediction (**Table 1**).

Hotspot prediction	**Properties Used**	**Availability**	**Technique**	**Reference**
FoldX	Energy-based	Tool & Server	Computational alanine scanning	(76)
Robetta	Energy-based	Server	Computational alanine scanning	(77)
PP_Site	Energy, structure-based	Tool	Simple algorithm	(78)
FTMap	Energy –based	Tool & Server	Probe based rigid body docking with fast Fourier transform correlation	(79, 80)
PCRPi, PCRPi-W	Energy, structure, evolutionary	Tool & Server	Bayesian Network	(81, 82)
Guhary & Chakrabarti	H-bonding, interface location	Tool	Simple algorithm	(83)
KFC2	ASA, various structural features	Server	Support Vector Machine (SVM)	(84)
MINERVA	Structure, sequence,	Tool	Descision Tree, SVM	(85)

	molecular interaction			
HSPred	Energy-based	Tool	SVM (Residue specific)	(86, 87)
Rajamani	Side chain ΔASA	Tool	Molecular Dynamics	(65)
Higa & Tozzi	Structural, evolutionary	Tool	SVM	(88)
Grosdidier & Recio	Energy-based (Docking)	Tool	Normalized interface propensity	(89)
HotPoint & HotSprint	Conservation, accessibility, residue propensity	Server & Database	Empirical formula	(46, 90)

Table 1. Summary of Hot Spot Prediction Methods

1.2.4 Current Computational Methods to predict hot spots

1.2.4.1 Energy-based Computational Methods

Computational alanine scanning estimates the energetic contribution of each residue to the total binding energy via virtual alanine scanning. The process for computational alanine scanning is briefly described as follows: Using a side chain repacking algorithm, mutation of each interface residue to alanine is performed. Then, a numeric energy function is used to evaluate the bound and unbound states of the original and mutant proteins. Finally, the change in the binding energy (ΔΔG) of each

mutation is calculated thermodynamically. This approach requires the complex structure as input and is best for providing an accurate estimation of free energy charges (91).

One such algorithm is FOLDEF, available as the FoldX web server. It was developed by Guerois et al., and was used to predict the energetic effect of mutations on protein complexes. FOLDEF was built on the FoldX complex energy function, which systematically truncates side chains between two proteins (or protein-peptide) to alanine, then calculates change in the binding energy after relaxation (76). The complex energy function contains terms for: implicit and explicit desolvation, van der Walls forces, hydrogen bonding, Coulombic electrostatics, changes in entropy, and dipole interactions. FOLDEF was trained on a database of 339 mutants and gave a global correlation of 0.83 and a standard deviation of 0.81kcal/mol when tested on a database of 1,030 mutants. FoldX predictions have a linear correlation with experimental binding energies, and the standard deviation of predictions from a linear regression of the data is 0.88 kcal/mol (92).

Similarly, Kortemme and Baker used a simple physical model (named Robetta) that includes various parameters including: Lennard Jones interactions, solvation interactions, packing interactions, implicit solvation and hydrogen bonding to calculate free energy (77). Similar to FoldX, the Robetta energy function was parameterized from protein stability data. Robetta mutates side chains to alanine and locally repacks the structure within 5Å of the mutant residue, while keeping the rest of the protein unchanged. The predicted changes in binding energies form the basis for hot spot

predictions (93). When tested on a database of 743 mutations in the interface of each protein (from the ASEdb), Robetta was able to correctly predict 79% of hot spot residues with a 1.0 kcal/mol cutoff, and an average error of 1.06 kcal/mol (93, 94). The implicit solvation model has two potential errors: Robetta cannot identify hot spots that are involved in water-mediated hydrogen bonds, and when waters can compensate for a mutation, it will incorrectly predict mutations as hot spots.

Gao *et al.* developed a structure-based method, called PP_SITE, based only on three properties: H-Bond, hydrophobic and van der Walls interactions. PP_SITE is developed based on POCKET, which is a module of the multi-purpose program LigBuilder (95). PP_SITE creates a box with regularly spaced grids to cover interfacial residues and uses probes for each structural property to screen those grids. The probes include a positively charged sp^3 nitrogen (ammonium cation), a negatively charged sp^2 oxygen (as in a carboxyl group), and a neutral sp^3 carbon atom (methane). For each grid, the scores are calculated and a grid label is assigned to it. This method then offers the ability to analyze the decomposition of the contributions of regions in the predicted hot spots, enabling comprehensive exploration of the properties of the protein-protein interface. PP_SITE was tested on 13 complexes with 250 alanine mutations on interfacial residues. For the 75 hot spot residues in this set, 66 were predicted correctly, giving an 88% success rate ($\Delta\Delta G > 1.5$ kcal/mol) (78).

FTMap (79) is a recent approach that uses docking results of small organic probes to discover hot spots. It was originally conceived as a computational equivalent of the experimental multiple solvent crystal structures (MSCS) technique, where a

target protein is co-crystallized in the presence of a diverse set of organic solvent probe molecules. It has been shown that molecular probe distribution on the protein surface allows for discovery and characterization of potential ligand interaction sites (96). The FTMap algorithm samples billions of positions from 16 small molecule probes (with varying hydrophobicity and hydrogen-bonding capacity) globally across the total protein surface using a fast Fourier transform correlation approach. The poses are evaluated with an empirical scoring function that includes van der Waals terms, electrostatic interaction energy, a cavity term to represent the effect of nonpolar enclosures, and a structure-based pair-wise interaction potential. The 2000 most favorable docked conformations are energetically minimized and clustered. The lowest energy clusters from different and overlapping probe types are clustered again into consensus sites; the largest consensus site is generally located at the most important sub-site of the protein binding site. FTMap is available as a publically-available server, where only the PDB file or a PDB code of the protein is required. The output shows the six lowest energy cluster representatives as well as the number of non-bonded and hydrogen bonds between the probes and each residue in the protein (**Figure 4**).

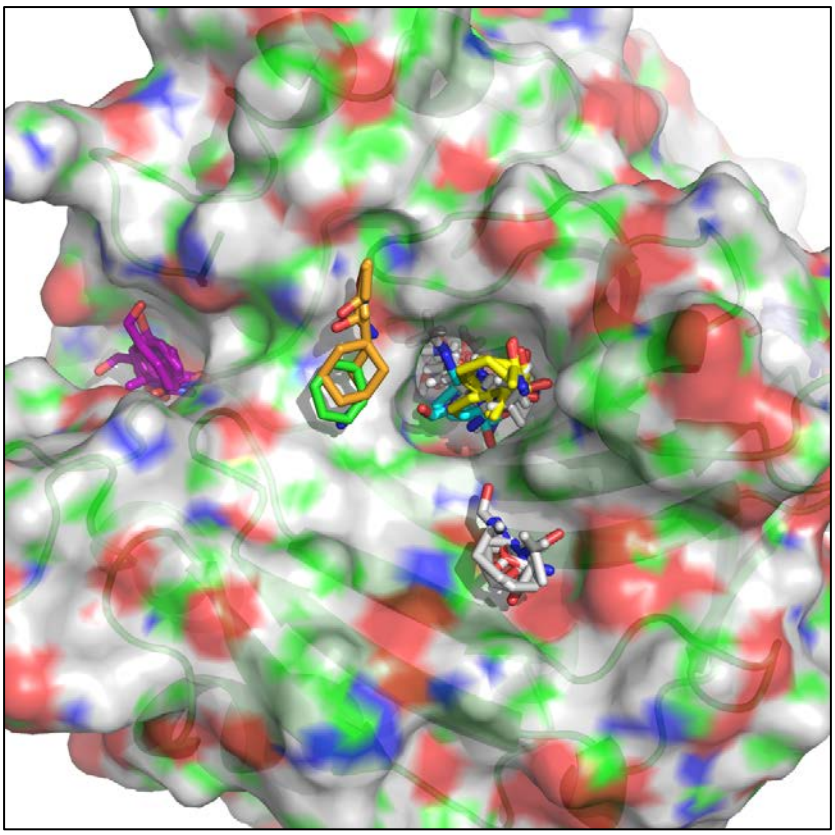


Figure 4. FTMap results for carbonic anhydrase I (PDB 2CAB) displayed in PyMOL. The pocket nearest to the center contains 19 docked probes, and is the binding site for drugs acetazolamide, methazolamide, and foscarnet

Another recent method called PCRPi (Presaging Critical Residues in Protein interfaces) is based on the integration of three main sources: energetic, structural, and evolutionary determinants by using Bayesian networks to unify them into a common probabilistic framework (81). Specifically, the attributes are: interaction engagement index, topographical index, four separate conservation indexes (ANCCON, ANC3DCON, CON, and 3DCON), and the BE index. The BE index is an *in silico* alanine scan using FoldX (76). This method has been developed as a web server (called PCRPi-W) where users can enter a PDB code or upload a complex, as well as select the type of Bayesian network architecture (naïve or expert) (82). Two different types of training sets are available for users: Ab+, which contains entire dataset of 636 interface

residues, including antigen-antibody complexes, and Ab-, which does not include these antigen-antibody complexes. Using a BID dataset derived in Tuncbag *et al*. (97), the PCRPi-W method recorded a precision of 0.79 and recall of 0.64 when compared in a head-to-head comparison with other computational methods (82). Also of note, the PCRPi-DB is the result of annotating and archiving the entire Protein DataBank (98) using PCRPi (99).

Finally, Guharoy and Chakrabarti used only the two criteria: hydrogen bonding across the interface, and location in the interface core to predict more than 80% of hot spots with experimental contribution of core residues to the energy of binding ($\Delta\Delta G$ > 2 kcal/mol) (83). An experimental database of 462 mutants from 28 interfaces was used, out of which 143 form hydrogen bonds. The average unsigned error (calculated as $|\Delta\Delta G_{calc} - \Delta\Delta G_{exp}|$) for all mutations was 1.04 kcal/mol. This technique is computationally inexpensive due to its small number of parameters, and is hence potentially applicable across a large set of complexes. One limitation of this technique is that coupling and indirect effects cannot be considered explicitly.

1.2.4.2 Molecular Dynamics-Based Methods

These types of methods are also energy-based, but use more intensive molecular dynamics computations to provide an atomic level of hot spot prediction. Therefore we grouped them separately. One particular study using MD techniques looked at anchor residues, which are residues that have limited mobility and strongly correlate with conserved hot spot residues (100). The study of anchor residues was conducted by

Rajamani *et al*. on 11 different proteins and concluded that these residues have side chains that experience the largest decrease in solvent-accessible surface area (>100 Å2) when in a complex (65). From their analysis of 39 complexes, they concluded that anchor residues have high structural conservation across different homologs. They also found that the anchor residue (primarily from the smaller of the two proteins) binds to a structurally constrained binding groove of the other protein that stabilizes the bound intermediate state, and the larger the buried surface area of the main anchor residue, the fewer secondary anchors are required for complex formation (65). This work was based on another group's nanosecond explicit solvent MD simulations of three complexes and the observation that key side chains frequently display the rotamer conformations of the complex before any receptor-ligand interaction (101).

Huo *et al*. applied a MM-PBSA (molecular mechanics-Poisson-Boltzmann surface area) approach to the human growth hormone-receptor complex (hGHpb). Here, protein-protein interfaces are probed with explicit molecular mechanical energies, continuum implicit solvation free energies, and approximations of vibrational entropy changes. All of these terms are then averaged over each conformation provided by the MD simulation. Using free energy calculations of snapshots equally spaced along a single dynamical trajectory, they were able to predict the experimental $\Delta\Delta G$ of binding with an average error of ±1kcal/mol for the alanine mutations of hydrophobic and polar/charged residues without buried salt bridges (102). The full structural and energetic consequences of a mutation of a bound and unbound state can be adequately described when the proper thermodynamic cycle is employed (102). This

comprehensive analysis requires repeated sampling of the mutant in both bound and unbound forms, which is computationally expensive when compared to other methods, especially on a large scale. To circumvent this potential computational roadblock, approximations to generate the mutant ensembles can be used which simulate first the un-mutated protein, then are post-processed to introduce mutations to alanine (64, 103).

1.2.4.3 Machine Learning-based Methods

The third method uses learning based methods (commonly called machine learning, or ML). This method is very novel, considers a diverse series of factors, and will be used in future drug discovery efforts. The KFC (Knowledge-based FADE and Contacts) uses a rule-based model created by a machine learning algorithm to elucidate structural patterns that are indicative of hot spots. The KFC model is actually a combination of two learned decision tree models: K-FADE uses the residue size and radial distribution of shape specificity and interface points calculated by Fast Atomic Density Evaluation (FADE) (104), and K-CON uses the residue's intermolecular atomic contacts, hydrogen bonds, interface points, and chemical types (105). These structural features return a binary answer as to whether a residue is a hot spot or not, and a confidence score with each prediction. Alone KFC predicted 58% of alanine scanning hot spots with precision of 49% and recall of 58% when a hot spot is associated with $\Delta\Delta G > 2$kcal/mol. When KFC was combined with Robetta's computational alanine scanning technique described previously, (the combined model called KFC+Robetta-Ala, or KFCA) (93), the predictive accuracy jumps to 72%. The

KFCA analysis was applied to two complexes: the calmodulin (CaM)/smooth muscle myosin light chain kinase (smMLCK) and to the bone morphogenetic protein-2 (BMP-2)/BMP receptor-type I (BMPR-IA) complex. Both complexes had strong correlation between KFC hot spot predictions and experimental mutations, and in the case of BMP2, KFC was able to highlight a region where alanine mutation of two key residues (Phe49 and Pro50) are more disruptive than the sum of the individual mutations (106). A more recent study also used the KFC method in the study of the interaction of the antibody, called 19D9D6, and the hepatitis C virus (HCV). Comparison of MD simulation results matched three residues that were predicted as hot spots in KFC that are important for the interaction between the virus and antibody (107). The KFC method is now available via an interactive, public web server, where users can submit complexes (via uploaded file, or providing the PDB code) and view results for each job, as well as upload scores from Robetta's alanine scanning, ConSurf sequence conservation, or experimental data (105).

The KFC method has been recently improved, with two new models trained using a support vector machine (SVM) (108): KFC2a and KFC2b. KFC2a is composed of eight features primarily related to solvent accessibility, interface position, packing density, and local plasticity. This KFC2a model showed a true positive rate of 0.85, outperforming HotPoint, KFC, Robetta, and FOLDEF, despite having a slightly higher false positive rate than the other models. KFC2b uses seven features (two of which are found in KFC2a), and boasts a higher specificity (84).

Cho *et al.* applied a SVM that initially incorporates 54 multifaceted features from structure, sequence, and molecular interaction. The model is named MINERVA, an acronym of MINE, Residue VAlue. From these 54 features, decision tree selection is performed to identify the best feature subset. Atomic packing density, relative surface area burial and weighted hydrophobicity are the top three features for picking hot spots that were selected from the decision tree, with the weighted atomic packing density displaying the highest effectiveness (85). Previous hot spot and conservation analysis has shown that local packing density around conserved residues and hot spots is higher than anticipated, and a good correlation exists between local packing density and experimental $\Delta\Delta G$ (109). MINERVA displays better sensitivity, specificity, precision and F1 score than Robetta, FOLDEF, and KFC when tested on two different training sets; the best performance occurred when the model was trained on the 2 kcal/mol set rather than the 1 kcal/mol set (85). Their work also revealed that residue conservation is not ideal for correctly identifying hot spots, and that hot spots are closely related to π- π interactions.

Another group that has successfully used SVM is Lise *et al.*, who previously described a hybrid method in which energy terms are used as input features of a SVM classifier and is called HSPred (86). Their method considers basic energy terms (van der Waals, H-bond, electrostatic and desolvation potentials, hydrogen bonds, and Coulomb electrostatics) calculated from the complex structure. In a later study, this model was improved by 10% in overall precision and recall by creating separate SVM classifier predictions specifically optimized for arginine or glutamic acid residues, as

these amino acid types did not perform well in the original model (87). The method is also available as a web server and is free to non-commercial users (87).

A set of 43 structural and evolutionary parameters were used with an SVM classifier in the work by Higa and Tozzi. These parameters can be grouped into the following types: amino acid type, evolutionary profile, conservation score, surface area, solvation energy, and geometry (88). When using the dataset compiled by Darnell *et al.*(106), this method had a performance of 60.4% (measured by F-Measure) that corresponded to a recall of 78.1% and a precision of 49.5% (88). This work is significant in that it does not require the complex; it can be used with only knowledge of the monomer. As a result, this method can be used when the interface region is unknown, and can prove useful for discovery of novel protein interfaces.

Another method that does not require prior knowledge of the complex structure is the work from Grosdidier and Recio, whereby the analysis of rigid-body docking ensembles provides normalized interface propensity (NIP) values. This NIP parameter is obtained from rigid-body docking simulations and calculates the propensity of a given residue to be located at the interface. It includes electrostatics and desolvation scoring in its calculations for predictions (89). The NIP method was developed from their previous findings that highly populated low-energy regions consistently corresponded to proven binding sites (110). Their method has up to a 80% positive predictive value, but it is not exhaustive and will miss hot spots that are not directly involved in the interface or arise from highly specific interactions (89).

While machine learning methods hold great promise, it is worth mentioning that simple empirical models can provide better predictive power. Tuncbag *et al.* presented an empirical model based on conservation, accessible surface area (ASA), and knowledge-based pairwise residue potentials of the interface residues This model provides 70% accuracy, higher than machine learning based methods when using 150 residues from ASEdb as the training set (97). Later, this method was incorporated into a web server, called HotPoint, where users can upload a protein complex and quickly visualize results (46). The relative residue solvent accessibilities are calculated in the complex and monomer states, and the solvent mediated potentials are taken from Keskin *et al* (111). If the following two qualifications for a residue are met: the relative ASA in complex is less than or equal to 20%, and the total pair potential is greater or equal to 18.0, then that residue is labeled a hot spot (46). Change of solvent accessible surface area (ΔASA) upon the formation of a complex is strongly related to the solvation energy, which has been shown previously to be one of the most significant factors involved in protein-protein binding (112). This technique was also used to correctly identify the hot spots on Mdm2. Another database named HotSprint was developed by the same group. It was the first database to exploit sequence conservation and incorporates solvent accessibility of residues. This method gives an accuracy of 76% that performs better than several machine-learning based methods (90).

1.3 The Ubiquitination Pathway

Sections reproduced from Future Medicinal Chemistry. (2015) 7(17), 2333-2350 with implicit permission of Future Science Group

1.3.1 Introduction

Ubiquitination, a step in the non-lysosomal degradation of proteins, is a crucial post-translational modification in eukaryotic organisms. Rapid and timely degradation of transcriptional regulators and other proteins by the ubiquitin-proteasome system (UPS) regulates a wide variety of cellular processes(113). Ubiquitination involves covalent attachment of ubiquitin, a small 8-kDa protein, to a substrate and results in recognition and shuttling of the substrate to the 26S proteasome complex for degradation (114). It is important to note that the ubiquitination process combined with the proteasome complex step is also referred to as the ubiquitin-proteasome system (UPS) or ubiquitin proteasome pathway (UPP).

The ubiquitination process is tightly controlled by three families of enzymes: ubiquitin-activating enzymes (E1s), ubiquitin-conjugating enzymes (E2s), and finally ubiquitin-protein enzymes (E3s). There exists two E1 enzymes with ubiquitin-activating capability : UBA1 being the primary E1 and the recently discovered UBA6 with unclear functions and uncharacterized regulations (115, 116) In contrast to the small number of E1s, there are approximately 40 E2s (117, 118) and 500-1,000 human E3 ligases, providing both specificity and versatility (119). The three steps of the ubiquitination process (**Figure 5**) have been reviewed previously (120, 121).

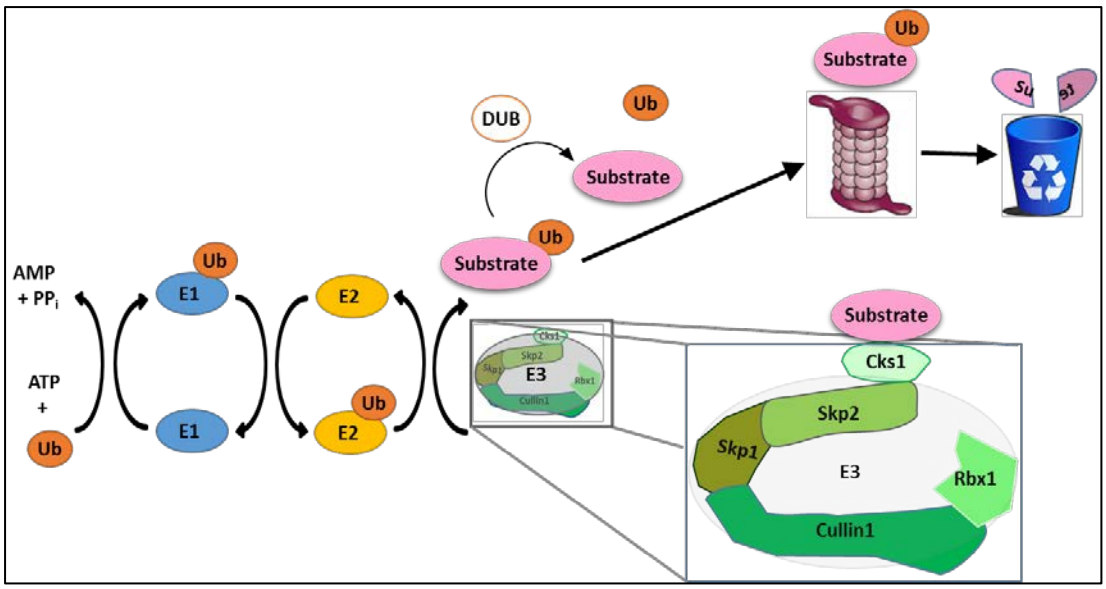


Figure 5. General structural outline of ubiquitin/proteasome pathway. E1s (Blue) are activated by ATP and bind to ubiquitin (orange). They then transfer ubiquitin to E2 (blue) and the E3 enzyme transfers ubiquitin to the substrate, which is then shuttled to the proteasome for degradation.

Briefly, the activation step requires binding of both ATP and ubiquitin and links the α-carboxyl group of the C-terminal glycine residue of ubiquitin to a cysteine residue on E1, and a thioester linkage is formed between the ubiquitin and E1. Then the E2 binds to both activated ubiquitin and the E1 enzyme and thus transfers the ubiquitin from E1 to the active site cysteine of the E2 via a trans(thio)esterification reaction. Finally, the E3 catalyzes the linking of ubiquitin to a lysine residue on the substrate. Repetitions of these sequential steps results in long chains of ubiquitin (polyubiquitin) on the protein to be degraded, and the specific lysine residue on ubiquitin used for linking (e.g., K48, K63, etc.) results in different topologies (122). Ubiquitination was originally described as a mechanism by which cells dispose of short-lived, damaged, or abnormal proteins, but more recent studies have revealed that it also plays a significant role in post translational modification. Ubiquitination can

result in the addition of a single ubiquitin moiety, called monoubiquitination, rather than polyubiquitination. Generally, polyubiquitination reactions are formed on the K48 residue, and this process tags substrates for proteasomal degradation and recycling (123). On the other hand, the K63-linked non-proteolytic ubiquitination spares proteins from degradation and regulates localization and activity of multiple kinases and pathways, such as PKB/Akt, TAK1, IKKγ/NEMO, TNFR, IRAK1, MLK3, IGF-1R, T cell receptor (TCR), NOD-like receptor (NLR), and RIG-I-like receptor pathways. This type of ubiquitination can cooperate with other linkage types to achieve the physiologically required output of a signaling pathway (122, 124, 125) and therefore has been crucially implicated in diverse biological processes including signal transduction, transcriptional regulation, growth response, innate immune response, and DNA repair and replication (124-126).

Mitotic cell cycle progression through each stage is tightly controlled by cyclin-dependent kinases (CDKs), and their relative interactions with members of the cyclin and CDK inhibitor (CKIs) families. The relative amounts of these signal factors fluctuate within and between each stage of the cell cycle occurs via the result of periodic proteolysis(127); the ubiquitin-proteasome scheme is responsible for the degradation of these mitotic regulatory proteins, which results in the control of their intracellular concentrations(128, 129).

1.3.2 Ubiquitination in Cancer

Ubiquitination can affect cancer development and progression in many ways. Both tumor suppressing and promoting pathways have elements that are tightly regulated by the process. One fundamental aspect of cancer is the deregulation of the cell cycle and checkpoint control (127). which is highly regulated through constant synthesis coupled to a particular timeframe of specific proteolysis of cyclins, cyclin-dependent kinases (CDKs) as well as CDK inhibitors (CKIs) executed by the UPS (130). Other cell cycle regulatory proteins that may contribute to cancer progression include cyclin E, p57, p130, FoxO1, c-Myc and E2F1 (131, 132).

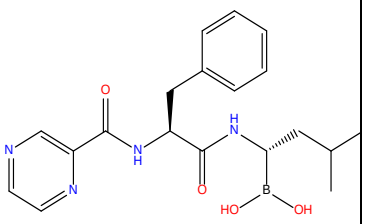
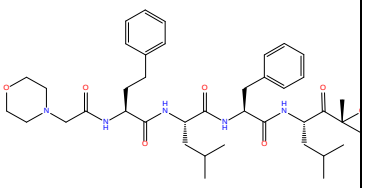
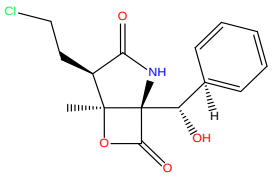
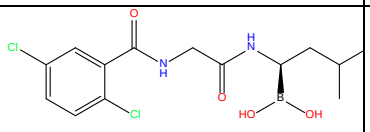
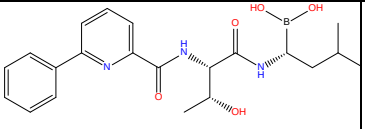
Another well-known example is the E3 ligase MDM2 which bind to the tumor suppressor protein p53 that is inactivated in more than 50% of human cancers. Also, mutations and alterations in ubiquitin ligases are found in a wide variety of tumor types and tremendously impact clinical outcomes (133-136).

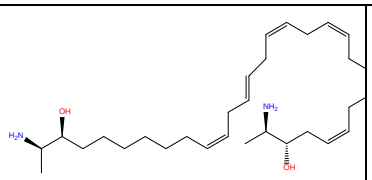
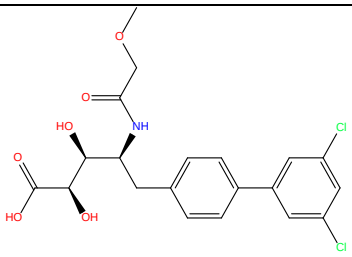
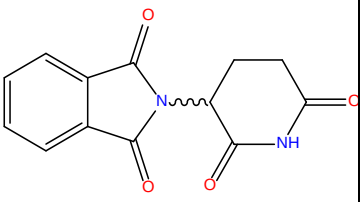
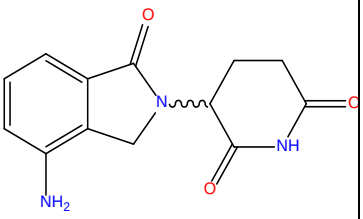
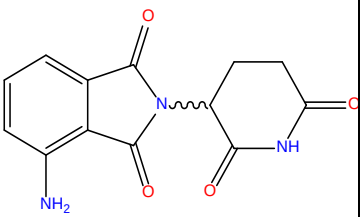
In addition to the above proteolytic polyubiquitination, which may contribute to cancer development, it is worth mentioning that monoubiquitination has unique effects on cancer as well (125). Monoubiquitin can serve as a recruitment signal to proteins that contain ubiquitin binding domains, and the functions of such non-proteolytic ubiquitination include, but not limited to: altered protein activity, subcellular localization, enzyme activation, DNA repair, chromatin dynamics (124, 137-139), and transcriptional regulation(140, 141). These facts underscore the importance of ubiquitination in tumorigenesis and the resulting interest as a clinical target.

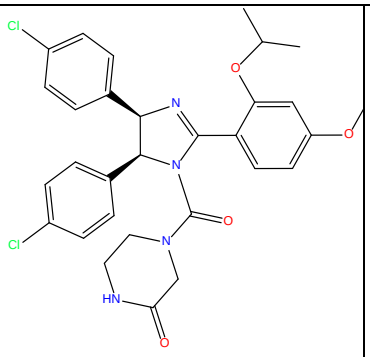
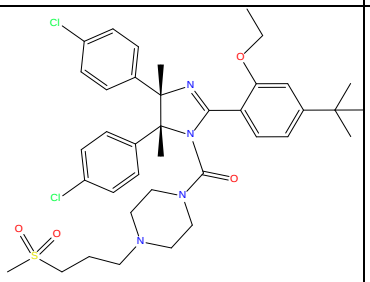
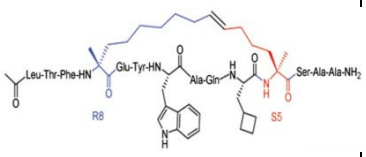
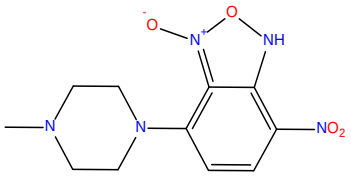
1.3 Existing Small Molecules Targeting Ubiquitin/Proteasome Pathway

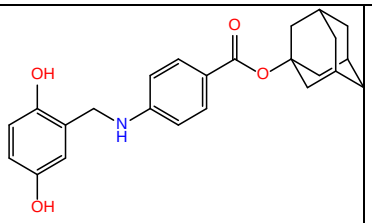
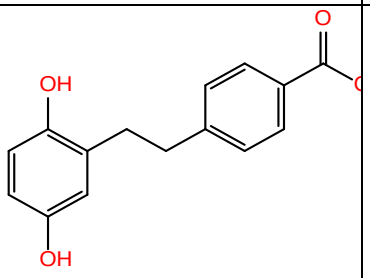
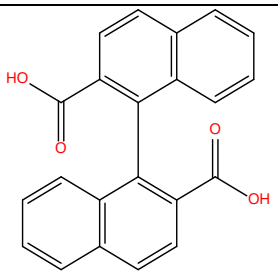
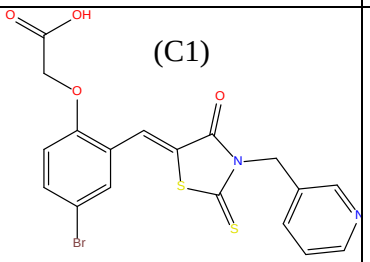
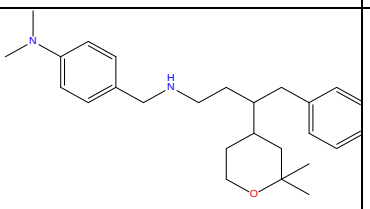
1.3.1 Targeting the Proteasome

The first therapeutic proteasome inhibitor tested in humans was bortezomib (Velcade), which was first synthesized in 1995, entered clinical trials in 1998, and approved by the FDA for use in multiple myeloma (MM) in 2003 (142) (**Table 2**). The clinical antitumor activity of bortezomib is well established as both a single-use agent and combination in the treatment of MM and other hematological malignancies. Bortezomib also exhibits efficacy in non-small cell lung cancer and pancreatic cancer (143, 144), and more recently it was expanded for use in patients with mantle cell lymphoma (MCL). High affinity and specificity of binding of bortezomib is achieved partly through the boron atom to the 26S proteasome's catalytic site (145-147). Consequently, proteasome inhibition alters the balance of all intracellular peptides, increasing those that require cleavage at acidic and hydrophobic sites and causing side effects such as neuropathy and autophagy in certain conditions (148-151). As is an issue with many cancer therapeutics, resistance can develop quickly and this occurs with bortezomib on average in about one year (140, 152, 153). Another notable proteasome inhibitor approved recently by the FDA is carfilzomib (marketed under the trade name Kyprolis), and is used for relapsed and refractory MM who have been previously treated with bortezomib (154, 155). Carfilzomib is also approved for MM and was derived from epoxomicin, a natural product that was shown to contain potent anti-inflammatory activity and proteasome inhibition (156).

Class	Drug Name	Target	Status	Structure	Ref.
Proteasome inhibitors	Bortezomib	20S Proteasome	FDA approved for MM and mantle cell lymphoma		(142)
	Carfilzomib	20S Proteasome	FDA approved for relapsed and refractory MM		(154, 155)
	Marizomib	20S Proteasome	Phase I		(157)
	Ixazomib (MLN-9708)	20S Proteasome	Phase I		(158)
	CEP-18770	20S Proteasome	Phase II		(159)

E2	Leucettamol A	Ubc13-Uev1A interaction	Preclinical/ Research		(160, 161)
	CC0651	Cdc34	Preclinical/ Research		(162)
E3 – immuno-modulators	Thalidomide	CRBN	FDA approved for treatment of MM		(163)
	Lenalidomide	CRBN	FDA approved for treatment of MM		(164)
	Pomalidomide	CRBN	FDA approved for refractory MM		(165)

E3 - p53 potentiators	Nutlin-3a	MDM2/ MDMX antagonist	Multiple Phase I trials		(166, 167)
	RG7112	MDM2/ MDMX	Phase I		(168)
	ATSP-7041	Dual inhibition of MDM2 and MDMX	Preclinical/ Research		(169, 170)
	NSC207895	Dual inhibition of MDM2 and MDMX	Preclinical/ Research		(171)

E3 - F-box protein antagonists	NSC689857	Skp2	Preclinical/ Research		(172)
	NSC681152	Skp2	Preclinical/ Research		(172)
	SCF-I2	Cdc4	Preclinical/ Research		(173)
	C1, C2, C16, C20	Skp2	Preclinical/ Research	(C1) 	(174)
	Compound A	Skp2	Preclinical/ Research		(175)

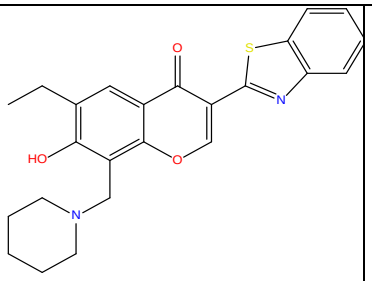
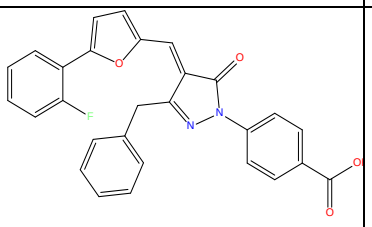
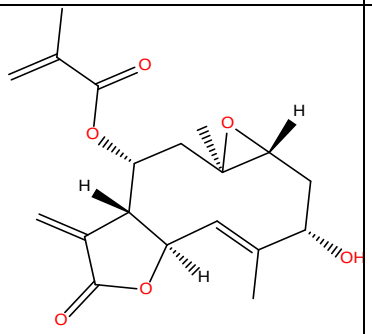
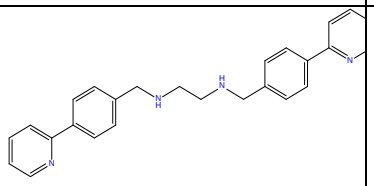
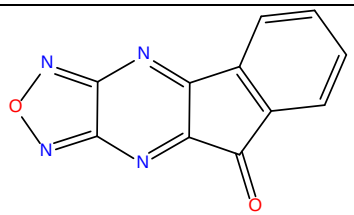
	Compound ZL-25	Skp2	Preclinical/ Research		(176)
	GS143	βTrCP	Preclinical/ Research		(177)
	Erioflorin	βTrCP	Preclinical/ Research		(178)
	BC-1215	Fbxo3	Preclinical/ Research		[77]
	SMER3	Met30	Preclinical/ Research		[75]

Table 2. Selected compounds targeting elements of the ubiquitin–proteasome system and their current status in the drug development pipeline.

Other next generation proteasome inhibitors are in clinical trials and have the potential to achieve a better therapeutic ratio and reduced probability of inherent and acquired resistance. An example is the structurally and pharmacologically unique marizomib (also called salinosporamide A) currently in Phase I trials (157). It is a natural product produced from marine bacteria, and its mechanism of action is unique in that it irreversibly and covalently modifies the active site threonine residues of the 20S proteasome (179, 180). Moreover, a Phase Ib clinical study of the combination of marizomib with vorinostat has just completed but the results have not yet been posted with non-small lung cancer patients (Clinical Trial NCT00667082).

Finally, two other reversible peptide boronic acid-based proteasome inhibitors in different stages of development are CEP-18770 and MLN-9708 (181, 182). CEP-18770 has been tested in Phase I trials for solid tumors and non-Hodgkin's lymphoma, and entered Phase II trials for relapsed and refractory MM by itself and in combination with lenalidomide (a thalidomide derivative) and dexamethasone (Clinical Trials NCT00572637, NCT01023880, and NCT01348919). It has shown *in vitro* antiangiogenic activity and potently represses RANKL-induced osteoclastogenesis, and is not cytotoxic to normal human epithelial cells and bone marrow-derived stromal cells (159). MLN-9708 (also called Ixazomib) has shown great promise and advanced to Phase III trials; it is also being considered in combination with lenalidomide and dexamethasone (Clinical Trial NCT01564537). In Phase II trials, while showing adverse side effects, it was generally well tolerated and the majority of patients (58%) had a very good partial response or better (158).

1.3.2 Targeting E2 Enzymes

While few therapies targeting the E2 enzymes are in development (relative to proteasome and E3 inhibition), there are some recent examples worth mentioning that are in preclinical stages (**Table 1**). Leucettamol A, which was isolated from the sea sponge *Leucetta aff. Microrhaphis*, was shown to inhibit the Ubc13-Uev1A interaction, blocking the E1-E2 complex formation (160). It was determined that, contrary to an earlier assessment that this was a racemic compound, it is in fact a chiral and optically active compound (161). Another marine sponge, *Lissodendryx fibrosa* was the source for manadosterols A and B which inhibited the same Ubc13-Uev1A interaction with higher potency than leucettamol A (183). CC0651 was found to allosterically inhibit the E2 enzyme Cdc34 and cause accumulation of p27 by inserting itself into a cryptic pocket distant from the catalytic site, causing displacement of secondary structural elements (162). However, targeting the E2 is still lacking sufficient specificity compared to E3 inhibitors as described in the following sections.

1.3.3 Targeting E3 Ligases

E3 ubiquitin ligases represent a diverse set of enzymes and provide the specificity of the ubiquitination reaction, and have significant roles in many different diseases, especially cancer. The E3 ligases are currently broadly classified into two major families from their structural motifs: HECT-type (Homologous to the E6-AP Carboxyl Terminus) and RING type (Really Interesting New Gene); RING domain

ligases are the most common (184). There are also two RING-finger derivative domains: U-box (UFD2-homology domain) and PHD (plant homeo domain) E3 types, but RING-finger types are the largest groups of E3 ligases.

E3 elements in the UPS are considered to be primarily responsible for the specific recognition of a large number of target proteins, acting as the substrate recognition component of the UPS pathway (114). This specificity is crucial when considering the prospect of designing drugs for the entire ubiquitination pathway, as they can be designed to target specific substrates of the E3 ligase without affecting other substrates. As a result, the E3 ubiquitin ligases are drug targets with the most potential for cancer therapies as they have fewer targets and offer a higher specificity of the system (185, 186). There are several notable E3 ligase targeting agents that are worthy of mentioning, and there have been many recent developments that show promising results in preclinical testing.

1.3.3.1 Thalidomide and its Derivatives

Thalidomide was originally developed as a sedative and an agent to cure hyperemesis gravidarum in pregnant women, and became infamous for its limb formation birth defects. Although banned after discovery of these effects in 1962, it had off-label use for patients of erythema nodosum leprosum (ENL); it was this use that led to clinical trials of its assessment and characterization of its immunomodulatory effects. Its primary target has been identified to be cereblon (CRBN), which forms an E3 ligase complex with damaged DNA binding protein 1

(DDB1) and Cul4A that is important for limb development, and binding of thalidomide to this complex appears to be the mechanism for the teratogenic effects in embryonic development (163). Two optimized second-generation derivatives of thalidomide, lenalidomide and pomalidomide (164, 165), are also used for MM by targeting CRBN and modulate its specificity as a substrate receptor, not by inhibition, rather by selective enhancement of the ubiquitination and degradation of Ikaros 1 and 3 (IKZF1 and IKZF3) zinc finger transcription factors (187). A crystal structure of lenalidomide in complex with CBRN shows the binding mode and that it binds to the substrate binding domain of CRBN and blocks ubiquitination (188-190).

1.3.3.2 MDM2/p53

The tumor suppressor protein p53 has been called the "guardian of the genome" due to its critical role in inducing cell cycle arrest and apoptosis in response to DNA damage. There exist over ten E3s associated with the regulation of p53, but the one with the most unquestionable importance has been MDM2 (murine double minute 2)(191). The p53-binding domain of MDM2 is at the N-terminus, and the RING domain in the C-terminus acts as the ubiquitin ligase to promote rapid degradation of p53 after its export from the nucleus (192-194). Inhibition of the MDM2/p53 interaction has been achieved by the small molecule nutlins, which have completed Phase I clinical trials and (166, 167). Nutlins are the general name given to compounds based on a *cis*-imidazoline scaffold, and their derivatives offer better activities but still retain elements of the core structure. Serdemetan was tested in a

Phase I trial and showed good p53 induction, but cardiac construction effects were observed (195-197). Nutlin-3a (a more promising agent) was optimized for better pharmacological properties to become RG7112 (also known as RO5045337); it stabilized p53 and activated the p53 pathway in cancer cells, and is being tested both as a monotherapy and in combination and some Phase I trials are still ongoing (NCT01677780) (168, 198). Nutlins and their derivatives are competitive inhibitors and structural mimics of p53 (via Phe19, Trp23 and Leu26) (166). MDMX (murine double minute X – also known as HDMX in humans, or also MDM4 or HDM4) shares significant homology with MDM2 and is also a negative regulator of p53, but nutlin-3 has shown decreased effectiveness in inhibiting MDMX-p53 interactions due to its differential binding (199). This highlights one limitation of MDM2-targeting agents such as nutlins: tumors with high MDMX and low MDM2 expression respond poorly to MDM2 inhibition (170). To combat this problem, a dual inhibitor of MDM2 and MDMX (ATSP-7041) was discovered and activates the p53 pathway *in vitro* and *in vivo* (169, 170). Another compound overcoming the limitation, NSC207895, was found to target MDMX specifically and acted additively with nutlin-3a to activate p53 and decrease cancer cell viability (171).

1.3.3.3 SCF E3 Ligases (Excluding Skp2)

The RING-finger type SCF (Skp1/cullin/F-box) E3 ligases are the largest family of E3 ubiquitin ligases, and they consist primarily of three core subunits: an F-box protein that contributes to the specificity of the SCF, Skp1 (S-phase-kinase-

associated protein-1) that acts as a bridging protein and binds to the F-box, and Cul1 (One of the seven family members in the Cullin family) that forms the major structural scaffold of the SCF. The majority of the substrates of SCF E3 ligases are involved in regulating cell cycle progression, gene transcription, DNA replication, and signal transduction (119, 200, 201). A growing body of evidence suggests that phenotypes such as genomic instability, uncontrolled proliferation, and cancer result from the dysregulation of these E3s (119), and more than a couple E3 ligases have been proposed to be drug targets and prognostic biomarkers in cancers such as melanoma (202). Their potential as drug targets has been only been uncovered very recently, and as the F-box proteins provide the specificity of the SCF complex, many small molecules have been designed to target specific members (**Table 1**).

SMER3 was discovered from a yeast-based screen of rapamycin enhancers, and blocks the SCF^{Met30} *in vivo* and *in vitro* but not the closely related SCF^{Cdc4}. SMER3 was also demonstrated to directly bind to Met30 and prevents degradation of Met4, an antiproliferative transcriptional activator (203). Small molecule screens identified SCF-I2, which inhibits the F-box protein Cdc4 in yeast but not its human ortholog Fbxw7 (173). Its specificity can be explained by an allosteric mechanism: it binds between two beta strands in the WD40 domain of Cdc4 that is 25Å away from the substrate binding site. While SCF-I2 failed to inhibit Cdc4 activity *in vivo,* it did show the potential of using allosteric inhibition of the WD40 domain.

Another example is BC-1215 which was recently synthesized to inhibit Fbxo3 and blocks degradation of another F-box protein (Fbx12), which in turn degrades the TNF

receptor-associated factor (TRAF) adaptor proteins that are responsible for cytokine secretion (204). This resulting TRAF inhibition by BC-1215 dampens NF-κB activation through the TNF signaling cascade (205).

βTrCP (b- transducin repeats-containing proteins) is an E3 ligase that binds to Skp1 and promotes degradation of a breadth of key regulatory protein elements in cancer biogenesis, including: Pro-caspase-3, WEE1, MCL1, p100, p105 and CD4. In most cases, it functions as an oncoprotein and has been found with upregulated mRNA levels in many cancer types (202, 206). GS143 and Erioflorin both block interaction between βTrCP and their targets: phospho-IκBα (nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha) and Pdcd4 (Programmed cell death protein 4), respectively (177, 178). Inhibition of βTrCP can result in tumor cell arrest at multiple points in the cell cycle at nanomolar concentrations and the inhibitors show promising results in vitro (207, 208)

1.4 Skp2 in the Ubiquitination Pathway and its Role in Cancer

1.4.1 Basic Information of Skp2 and structural features

One well characterized E3 ligase is the SCF (Skp1/Cul1/F-box) complex, which are RING-type (Really Interesting New Gene) E3s and these target a broad range of proteins involved in the progression of the cell cycle, transcription, and signal transduction proteins (209, 210). The Cul1 protein joins with Rbx1 to form the RING domain, and this forms the catalytic core complex that recruits the E2 portion, and the Skp1 acts as an adapter protein that is responsible to bind to the variable F-

box protein; this F-box protein is the variable portion of the SCF complex that binds to the substrate, and its consensus sequence is about 50 residues long with the most commonly repeated residue being leucine.

The most well characterized mammalian F-box protein is Skp2 (S-phase kinase-associated protein 2; also known as FBXL1, FBL1, and p45) (**Figure 6**). Skp2 acts as the substrate-recruiting element of the SCF and together with Skp1, Cul1, and Cks1 forms an E3 ligase to form the functional SCF complex. The most prominent and arguably significant protein that is ubiquitinated by Skp2 is the CDK p27. The protein p27 is periodically expressed in the cell cycle. Its specific phase of high expression are at the G_1 phase and it decreases at the S phase. As the levels of p27 mRNA remain stable throughout the cell cycle, it is widely considered that its protein levels are controlled means of a posttranscriptional mechanism. Skp2/SCF primarily targets p27 in the nucleus in the S and G_2 phases. As p27 is ubiquitinated by Skp2, there is an inverse relationship between Skp2 expression and p27 levels in multiple tumor types as well as normal tissues (211-213).

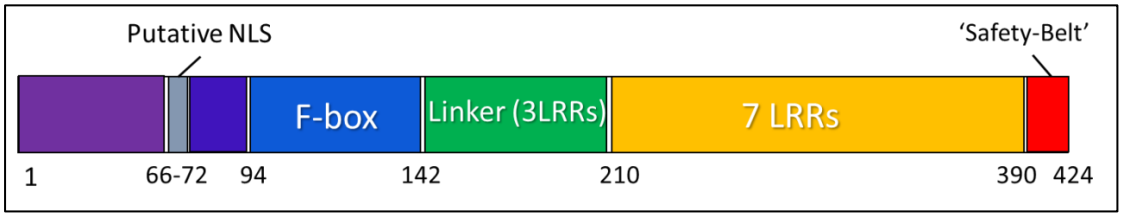


Figure 6. Linear representation of full length sequence of human Skp2 showing significant domains and features.

The Skp2 protein is 424 residues in length, with the F-box domain lying closer to the N-terminal region at the 94-140 position, and the C-terminal region forming a concave surface comprised of ten leucine-rich repeats (LRR) (**Figure 7**) Skp2 also

contains a putative nuclear localization signal (NLS) within the region where Skp2 phosphorylation occurs at by Akt at Ser72, triggering Skp2 cytosolic localization (214). The three LRRs closest to the F-box are non-canonical, while the other seven are canonical. After the tenth LRR, the ~30 residue C-terminal tail turns back towards the first LRR, forming what has been referred to as a 'safety-belt' that might aid to pin down substrates into the concave surface formed by the LRRs (215) (**Figure 8**). It is this C-terminal tail and the LRRs farthest from the F-box (repeats 6-10) that bind the Cks1 accessory protein, which in turn is required for the recognition and ubiquitination of the $p27^{Kip1}$ (216, 217).

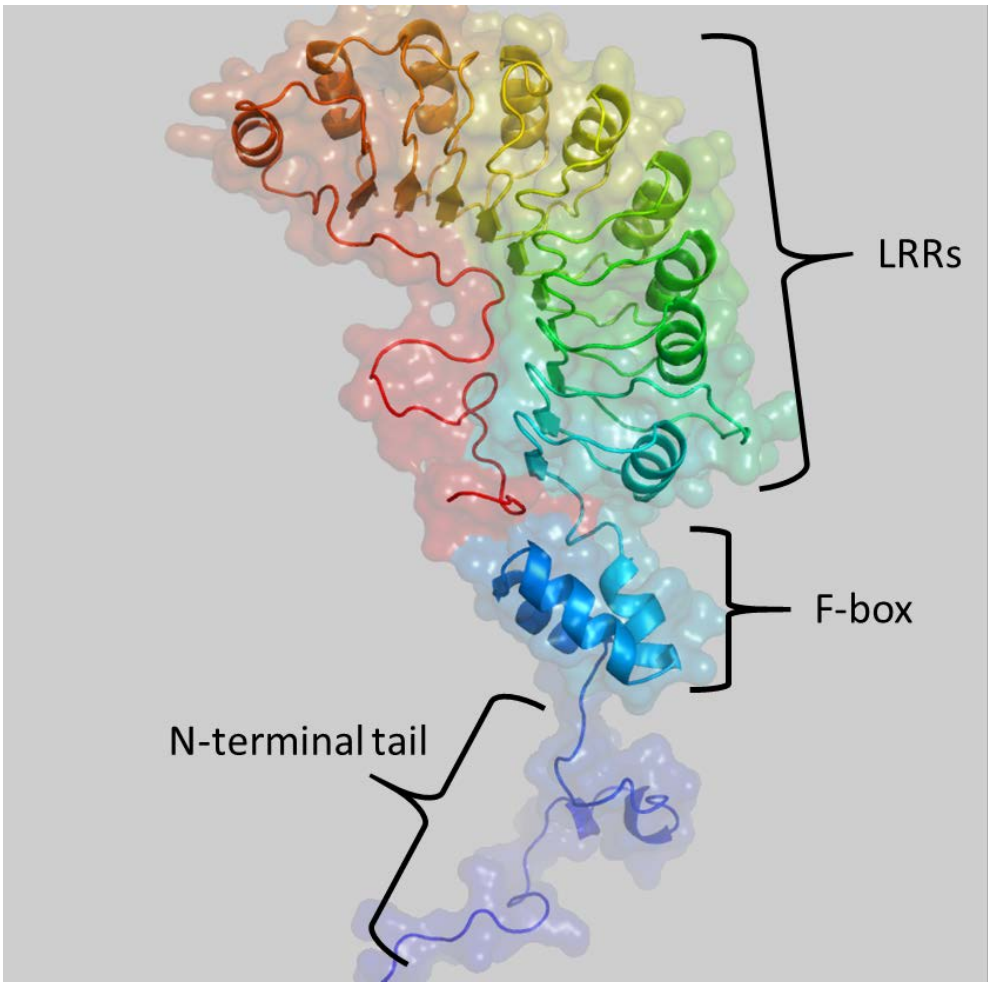


Figure 7. General structural of Skp2 colored from N-terminus (purple) to C-terminus (red). 2AST structure with N-terminal purple region is a homology model and prediction of tail in unbound Skp2.

1.4.2 Functions of Skp2 Interacting proteins and Roles in Cancer

Skp2, due to its role in degrading a significant cell cycle control element, is considered an oncogene and has been implicated in many different and diverse cancers, including: lung carcinomas (218), neuroblastoma and gliomas (219, 220), renal cell carcinoma (221), colon carcinoma and rectal cancer (222, 223), gallbladder carcinoma (224), gastrointestinal stromal tumors (225), esophageal and oral squamous cell carcinoma (226, 227), osteosarcoma (228), and cervical (229). In breast cancer, it has been shown that Skp2 expression is associated with acquisition of epithelial-mesenchymal transition (EMT) in paclitaxel-resistant breast cancer cells (230), and is a potential biomarker for Luminal A breast cancer (231). One recent meta-analysis of Skp2 expression that represented 5,514 patients (a large amount of which were Asian) showed Skp2 overexpression is associated with poor overall survival as well as disease-free survival/relapse-free survival (DFS/RFS) in all patients as well as in the analysis of subgroups. Analysis of multiple datasets indicate that these effects are due to amplification, and not mutation/deletion, of Skp2. Multiple studies support this notion, showing that copy number gain is the contributing mechanism of Skp2 overexpression (232) (**Figure 8**).

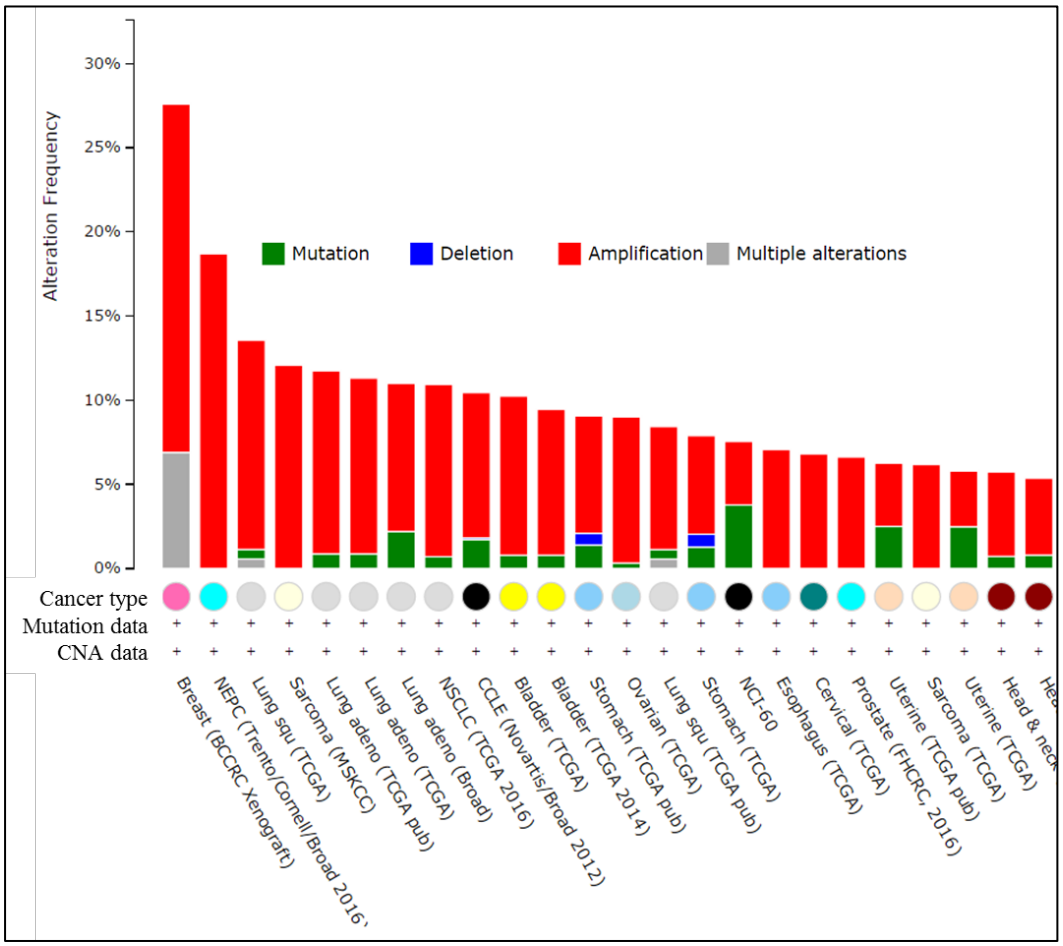


Figure 8. Skp2 is commonly amplified and not mutated in many cancers. Source: cBioPortal.com(233)

Recently some novel biological functions of Skp2 were identified and it can activate Akt through non-proteolytic K63-linked ubiquitination for glycolysis (via activating Glut1 expression) (234). A connection to tumor epithelial-mesenchymal transition (EMT) has been shown to function in Skp2 also via the increased Akt1 phosphorylation and the accumulation of c-Myc during EMT by TGF-β1 signaling to Skp2 via Akt1. Also, a recent report shows that Skp2 activates LKB1 (via K-63 ubiquitination) for cancer cell survival under energy stress via oncogenic Ras upstream of Skp2 (235).

It is well understood that for p27 to have its cyclin-dependent kinase inhibition (CKI) activity in negatively regulate the cell cycle progression, it must be present in the nucleus, and its activity is not present in the cytosol. Researchers have shown that the inverse correlation between Skp2 and p27 is clinically associated with poorer patient survival and increasing malignancy (213). Also significant is the fact that multiple cancers show that p27 is frequently lost or mislocalized (and not often mutated) to the cytoplasm, including cancers such as endometrial carcinoma (EndoCa), renal cell carcinoma (RCC) and breast cancer (236). In fact mutations in the *p27* gene across all cancers are exceedingly rare, at a rate of only 3% (237). The ability of p27 to block cell cycle progression is regulated predominantly by post-translational modifications; these determine the subcellular localization and the stability of p27. The ubiquitination of p27 is readily and quickly achieved in the nucleus specifically by the Skp2 SCF complex when p27 phosphorylated on Thr187, but when it is phosphorylated on Ser10, it is able to bind to CRM1 and be exported to the cytoplasm, where there it is targeted for degradation by another E3 ligase: the Cul1-KPC1/2 complex (238, 239).

The localization of p27 has multiple, startling effects on a cell. In the cytoplasm, p27 can repress RhoA, which leads to: diminished focal adhesion formation, reorganization of the actin cytoskeleton, and promotion of cell motility in addition to promoting autophagy and inhibiting apoptosis (240). Additionally, loss of nuclear p27 results in increased CDK2 activity and this increase results in phosphorylation of the estrogen receptor (ER), thereby elevating ER activity and

promoting growth of hormone-dependent tumors. These facts taken together indicate that p27 is a tumor suppressor in the nucleus, but an oncogene in the cytoplasm.

Therefore, there are two potential targets on Skp2 that might be exploited: the first is the Skp1/Skp2 interface. It could be disrupted, removing the rate limiting component (Skp2) from the SCF. The other is the p27 binding site. The specificity for p27 binding to the Skp2 SCF complex in the nucleus is made available by a pocket formed by Skp2 and an accessory protein named cyclin-dependent kinase subunit 1 (Cks1). This pocket provides an interface on which to target specifically that could result in the rescue of p27.

Chapter 2: Skp2 Hot Spots and Virtual Screening for Inhibitors of Skp2

Skp2 is clearly an oncogene and acts as the rate-limiting component of the SCF complex, and provides the specificity towards the substrate (most notably p27). Despite its role in transferring ubiquitin to a substrate, it has no prototypical "catalytic residue" that could be targeted like other enzymes. Skp2 also acts as an intermediate adapter between Skp1 and the accessory protein Cks1, and the PPI interfaces formed between both proteins is large, so conventional virtual screening techniques are likely to fail. To circumvent this problem and to design small molecules towards Skp2, we must target areas of the PPI interfaces in exactly the right region to disrupt the proteins, and the consideration of hot spots have the potential to find this region.

The main thrust of this study is to combine the theory of targeting PPIs (focusing on the theory of hot spots) and use a variety of computational techniques to predict the hot spots that exist on Skp2. If the hot spots cluster in a specific region that appears to be druggable (i.e., a pocket or groove), than this region will be selected as the center region on which to conduct a virtual screen. A successful screen in which a small molecule can disrupt the PPI and result in Skp2 unable to bind to the remaining members of the SCF complex. This chapter describes how the hot spot analysis was carried out, how the hot spots were selected, how the screens were conducted, and how a small list of molecules were selected for biological testing. Finally, this chapter shows the biological effects of the top hits.

2.1 Materials and Methods: Development and Application

2.1.1 Selection of Compound Libraries for Virtual Screening

There were multiple iterations and multiple sources that were used for the compound libraries. For the ligand set used in the initial Skp1/Skp2 virtual screens, we selected the chemical library DiverSet (241) from ChemBridge. DiverSet has a unique set of 50,080 handcrafted compounds that are derived from the Express-Pick collection of 450,000+ compounds. The majority of these compounds have good drug-like and lead-like molecules, with most compounds having molecular weights below 500 and logP values less than 5, in keeping with Lipinski rules of 5 (**Figure 9**).

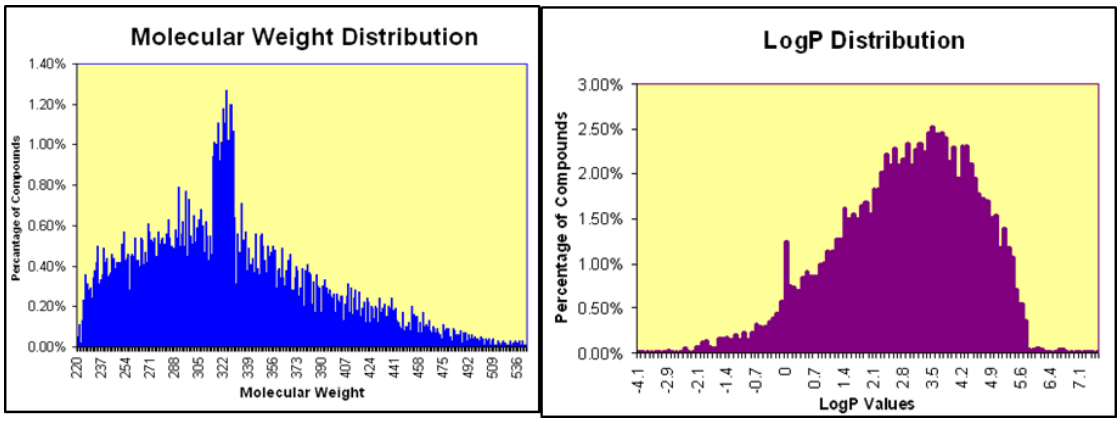


Figure 9. Distribution of molecular weight and logP of Diverset compound library used in Skp1/Skp2 screen

For the Cks1/Skp2/p27 screening, we used a chemical library that was first filtered so that it only contained drug-like and lead-like compounds then additional filters using our in-house protocol to remove PAINS (pan-assay interference compounds) to keep compounds that had logP within 1.0 to 4.0 and molecular weights between 100 and 400 were applied. This lead-molecule like library of

approximately 52 thousand compounds was then used as the basis for the virtual screening using our in-house HiPCDock program.

2.1.2 Virtual Screening Methodology for Cks1/Skp2/p27 binding site

Sections of this methodology were adapted from the following publication: "Targeting TRAF6 for Cancer Therapeutical Development", published by the University Of Texas Graduate School Of Biomedical Sciences at Houston in August 2012.

For the *in silico* high-throughput (HTS) virtual screening on the Cks1/Skp2/p27 binding site, the 2AST structure was obtained from the PDB (242), and the structure was prepared using the MOE software package from CCG. It is known that Cks1 is required for Skp2-mediated ubiquitination of p27, so the Cks1 protein was included along with p27 and Skp2 (216). Previous studies of p27 and its interactions demonstrated that mutation of Glu185 dramatically reduced the binding of p27 to the complex while T187A mutation completely abolished the phosphorylation of p27 (242, 243). We also found that in this structure, p27 shows two potential regions that were pocket-like that corresponded to these two residues. To further define the key residues to be included for virtual screening, we performed rigorous "hot spot" predictions using routine approaches in an ensemble manner(244), combined with molecular visualization of multiple SCF complex structures(206, 242, 243, 245).

Instead of performing one screen that covered both pockets, we performed three screens; two separate and exhaustive jobs were carried out on only the two individual pockets, and a final larger screen that covered both pockets together. The resulting hits from these screens were biased towards those that were present in both pockets. However, the majority of the top selected hits were found to occupy the Glu185 pocket. The top performing results from GOLD were visualized and examined for their ability to cohabitate both of these regions.

The resulting top 5,000 hits (based on docking scores and docking poses) were then docked using GOLD by CCDC. The top scoring 500 ligands (by GoldScore) were visualized individually (using PyMOL) in their docked conformation with the original submitted Skp2 2AST structure. Three main criterions were considered in the final selection of 500 hits. The first criterion was the degree to which the ligands' conformation occupied the interaction site between p27 and Skp2/Cks1. Second, the conformation, binding affinity, and resulting "fit" within the p27 binding site compared to the corresponding Skp2 residues was considered. Each ligand, given its orientation in the docked position, was evaluated on its ability to potentially block the binding of p27 protein by considering hydrogen bonds, electrostatic and aromatic interactions. The third and final consideration was the distinctiveness in the chemotypes when compared to other ligands. We placed the ligands into "classes" based upon the individuality of their underlying scaffold and how this scaffold was oriented in the binding site. Ligands that shared a similar scaffold and orientation were considered to be in the same class, and would then share the same chemotypes.

Also for the selected 500 hits, we assigned a “priority” to each molecule which was an aggregate of the second and third factors (conformation in the pocket and distinctiveness). This priority was defined as: low, medium or high. The priority then became similar to a degree of confidence for each ligand. Compounds with “high” priority were regarded as most likely to yield good biological activity, and should be tested over the other low and medium confidence ligands. Eventually from the 500 hits, we selected 123 unique compounds with high priority that were recommended to be tested for biological activity. In addition to these 123 unique compounds, based on the core scaffold of compound 405, a search of structurally similar compounds (also called a SAR by catalog) was carried out and identified 41 compounds that were analogues of compound 405.

2.1.3 Virtual Screening Methodology for Skp1/Skp2 binding site

The high-quality and most comprehensive crystal structure of Skp2 was obtained from the Protein Data Bank(246) (2AST) and the B chain (Skp2) was analyzed in complex with the A chain (Skp1)(242). The N-terminal tail has not been crystalized on this structure, but the rest of the protein (the F-box, the LRR domain, and the seatbelt region) is present in this structure. In fact, the N-terminus has never been crystallized on Skp2. Water molecules along with the C and D chains were removed from this 2AST crystal structure via PyMOL (247). From the literature, the 19 residues on Skp2 in contact with Skp1 formed the basis for detecting druggable sites in the F-box region of Skp2 (243). The first region (pocket 1) was close to the N-

terminus of Skp2, and the second region (pocket 2) had close proximity to the corresponding C-terminus of Skp1. The software MOE(248) and its site finder function validated the second pocket region on Skp2 within the F-box for which to dock small molecules to. Since the two regions (referred to as pocket 1 and pocket 2 henceforth) appear to have distinct differences in their surface area, size and location on Skp2, the two pockets were treated as separate virtual screens, each with a unique center and size.

The broad methodology for the virtual HTS on this region of Skp2 was similar to that for the Skp2/Cks1/p27 binding site. The docking program used, HiPCDock (249), is based upon DOVIS (250). The program considers a 3-dimensional cubic region with user defined coordinates, and the centers of these cubes were placed in the binding sites for each docking study. Pocket1 was centered at AutoDock coordinates 1.134, -114.378, -1.751 (x,y,z) and grid points of 80 in each dimension. Pocket2 had a center of -1.962, -100.846, 10.708, grid points of 48 in each dimension. Both virtual screens had genetic algorithm parameters of 100 runs, a population size of 100, and rate of crossover at 0.8. Further details (configuration files) of HiPCDock and the docking parameters used are available in the appendix.

After the initial docking of DiverSet, the 2000 compounds with the best HiPCDock scores for each pocket were then docked again into the same receptor (Skp2) using the GOLD (Genetic Optimization for Ligand Docking) software suite(251) to further narrow down the top binders using a scoring method different from HiPCDock. GoldScore was chosen as the fitness function, and a radius of 10Å

was used. The centroids were defined in the same manner as the HiPCDock, and a cavity file was generated from the selection. Further specific parameters used in the GOLD docking are available upon request.

The resulting top scoring 200 ligands from each pocket on Skp2 with the highest GOLD fitness values were then structurally clustered based on structural MACCS (Molecular ACCess System) fingerprints using MOE fingerprint clustering. The metric used was the Tanimoto coefficient, and the similarity and overlap were set to 60%. After clustering based on these fingerprints, the cluster center ligands from a given cluster are retained. The resulting 120 ligands from this clustering (70 ligands for pocket1, 50 ligands for pocket2) were then individually visualized (using PyMOL) in their docked conformation. Three main criterions were considered in the final selection process. The first criterion was the degree to which the ligands' conformation occupied the interaction site between Skp1 and Skp2. If the conformation primarily existed outside of the interaction site, it was no longer considered a candidate. Second, the conformation, binding affinity, and resulting "fit" within the Skp2 binding site compared to the corresponding Skp1 residues was considered. Each ligand, given its orientation in the docked position, was evaluated on its ability to potentially block the binding of the Skp1 residues. Along with these potential blocking effects, we also considered the binding affinity of the ligand in its pocket. This included qualities such as hydrogen bonds, static and aromatic effects. The third and final consideration was the distinctiveness in the molecular scaffold when compared to other ligands for each pocket. We placed the ligands into "classes"

based upon the individuality of their underlying scaffold and how this scaffold was oriented in the binding site. Ligands that shared a similar scaffold and orientation were considered to be in the same class, and would then share the same class number. We then assigned a "priority" to each molecule which was an aggregate of the second and third factors (conformation in the pocket and distinctiveness). This priority was listed as: low, medium or high. The priority then became similar to a degree of confidence for each ligand. Compounds with "high" priority were regarded as most likely to yield good biological activity, and should be tested over the other low and medium confidence ligands. The result from this visual selection process was that 57 ligands in total were recommended to be tested for biological activity; 41 total ligands were chosen as compounds to be tested for pocket1 with 16 compounds listed as high priority. For pocket2, 16 total compounds were selected, and 11 of these were high priority. Only the high priority compounds were actually tested biologically, and two of the ligands (7957666 and 9040817) were included in both of the pocket 1/pocket 2 lists; this lead to a final total of 25 compounds ordered for biological testing.

2.1.4 Biological testing by Collaborators

Cell culture and reagents

293T, PC3, A549, H460, H1299, Hep3B and U2OS cells were cultured in DMEM containing 10% FBS while LNCaP and H3255 were cultured in RPMI-1640 medium containing 10% FBS. (His)6-ubiquitin, GST-Akt, and HA-Akt constructs was described previously. (252) Flag-Skp2 was a gift from Dr. W. Wei. GST-Skp2

Trp97Ala (W97A) and GST-Skp2 Asp98Ala (D98A) mutants were generated in pGEXSkp2ΔN-Skp1Δ, a gift from Dr. B.A. Schulman, by site-directed mutagenesis assay according to manufacturers' standard procedures (Stratagene). ZL-25 and its analytical data including 1H, 13C NMR and LC-MS/MS analysis were obtained from ChemBridge. LY294002 was purchased from Cell signaling.

LC-MS/MS analysis of the in vitro binding between Skp2 and Skp2 inhibitor.

GST-Skp2 W97A and GST-Skp2 D98A mutants used for demonstrating #25-Skp2 interaction were generated in pGEXSkp2ΔN-Skp1Δ bicistronic co-expression constructs. Subsequently, the GST-Skp2 WT, GST-Skp2 W97A and GST-Skp2 D98A mutants were co-expressed, purified and formed complex with Skp1 (Δ38–44 aa.) as described previously (253). The purified proteins were incubated with ZL-25 for overnight at 4°C, and the GST tags were thrombin-cleaved and pulled-down by glutathione sepharose beads (Invitrogen). The untagged protein samples were pH-adjusted by adding 50mM ammonium bicarbonate and digested by adding 200 ng modified trypsin (sequencing grade, Promega) for 18hrs at 37°C. The resulting peptides and compound were analyzed by LC-MS/MS on an Orbitrap-Elite mass spectrometer (Thermo Scientific, Waltham MA). Proteins were identified by database searching of the fragment spectra against the SwissProt (EBI) protein database using Mascot (v 2.3, Matrix Science, London, UK). Typical search settings were: mass tolerances, 10 ppm precursor, 0.8d fragments; variable modifications, methionine sulfoxide, pyro-glutamate formation; up to 2 missed cleavages. The extracted ion

chromatograms (XICs) for the bound compounds were examined and peak heights were analyzed and quantified by using Qual Browser (Thermo).

2.2 Results: Predicted Hot Spot Residues for Virtual Screening

2.2.1 Hot Spots on Skp1/Skp2 interface

The Skp1-Skp2 interface is interdigitated, with alternate structural elements from protein interlocking with the other. The middle F-box domain of Skp2 and Skp1 C-terminal helix (H8 helix) are sandwiched between the other Skp1 helices on one side and the first LRR and C-terminal safety-belt on the other side. This interface has been described as having a variable and a core interface from the F-box perspective. In the core interface, the F-box packs with the H5, H6, and H7 helices of Skp1. It is referred to as the core interface because it accounts for two-thirds of the surface area buried, (2,050 $Å^2$) and contains residues conserved in all Skp1 and F-box protein members(243). The variable interface was defined as the opposite face of the F-box packing with the H8 helix of Skp1 and also with the first LRR and C-terminal end of the safety-belt (an area of 930$Å^2$). The structural elements in the variable interface are conserved in Skp1 and Skp2 orthologues, but not in other F-box and Skp1 family members, particularly in the H8 helix (243). This is not to say that the variable region does not contribute to binding. An in vitro assay that compared dissociation of the Skp1-Skp2 complex with that of a F-box-Skp1ΔH8 complex, which has a core interface structure similar to the original but lacks all of the elements unique to the

variable interface, showed half-lives of greater than 9 hours and less than 30 minutes, respectively(243).

The SiteFinder program in MOE (Chemical computing group) was used on the 2AST crystal structure, this program does not consider hot spots but uses simple proximity-based rules of connectivity. Also, the HotPoint, FTMap, as well as KFC2A and KFC2B models were used on the same structures. For each of these jobs, the 2AST structure was stripped of all other proteins/solvent molecules/waters, leaving only the Skp2 protein. Default parameters were used for the HotPoint models, and KFC2 was submitted the refined PDB structure, as the entry of the PDB code on the web server gave an error. KFC predictions were showing highly inconsistent results with the HotPoint and MOE site finder predictions, including suggestions on both Skp1 and Skp2 where residues had their side chains slightly buried and predictions of residues with smaller sidechains where it was not obvious as to how these residues interactions between Skp1 and Skp2 were possible. The KFC results were not discarded, just given less weight in the final ensemble consideration. Also, FTMap's prediction was biased towards pockets of the Cks1 binding site, but did find regions that were closer to Skp1. The top ranked cluster centers (the areas with the most docked pharmacophores) were corresponding to the Cks1 binding site (**Figure 10**). The top ranked cluster centers that were not in the Cks1 binding site region (in the LRR domain) actually corresponded to areas where HotPoint and SiteFinder both predicted there were concentrations of hot spot residues. These regions were close to

the Skp1-facing H1 helix on Skp2, and the region just behind the H2 and H3 helixes of Skp2, facing the beginning of the LRR domain.

To bolster the predictions, structural analysis was carried out on the contacts between Skp1 and Skp2 based on distance and hydrophobic interactions. Overall, HotPoint faired the best in predicting the contacting residues, but was biased towards residues that were larger and bulkier.

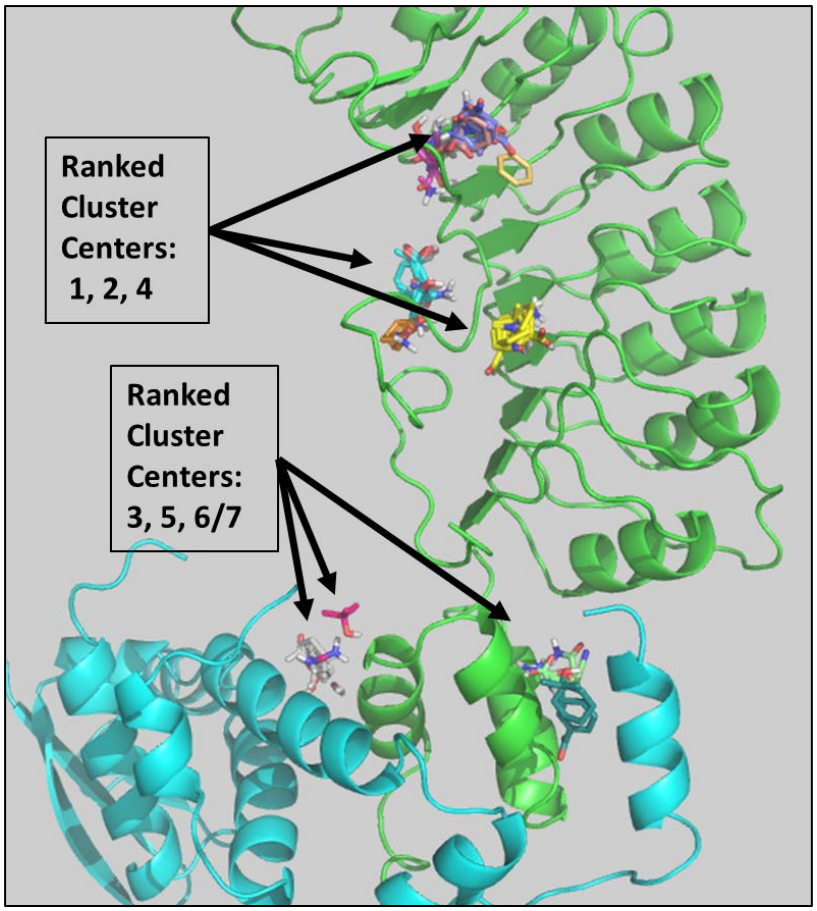


Figure 10. FTMap results of Skp1/Skp2 (blue/green) hot spot predictions. Top scoring cluster centers were in the Cks1 binding site, but centers 3, 5, as well as 6/7 showed good consensus to other hot spot predictions and residues contacting Skp2.

Based on visualization analysis as well as literature reports(242) there are 19 residues on Skp2 in contact with Skp1 and they form the basis for detecting druggable sites in the F-box region of Skp2 (206, 242, 243). The consensus of the hot spot

prediction tools as well as the SiteFinder results suggested we group Skp1/Skp2 contact sites into two pocket-like and distinct regions (**Table 3**).

Residue Number	Name	Chain	RelCompASA	RelMonomerASA	Potential	Prediction
95	V	B	62.3	115.13	12.39	
96	S	B	58.17	73.11	2.13	
97	W	B	3.16	61.76	32.59	Hotspot
99	S	B	69.61	80.8	8.13	
100	L	B	0.13	23.79	41.47	Hotspot
101	P	B	30.39	44.78	7.19	
104	L	B	2.83	49.91	31.55	Hotspot
107	G	B	30.84	48.2	10.65	
108	I	B	0.87	47.88	39.69	Hotspot
111	C	B	30.5	82.75	20.49	
112	L	B	2.32	24.76	32.47	Hotspot
114	L	B	0.08	5.13	61.89	Hotspot
116	E	B	16.96	37.81	8.76	
118	L	B	11.62	54.47	37.19	Hotspot
119	K	B	25.6	66.17	4.7	
121	S	B	6.27	19.42	13.14	
122	G	B	25.04	87.07	5.36	
123	V	B	2.78	60.03	17.75	
124	C	B	12.56	43.95	12.65	
125	K	B	53.92	91.3	8	
127	W	B	4.73	26.39	33.1	Hotspot
128	Y	B	35.12	65.45	18.64	
142	L	B	0.58	4.03	47.82	Hotspot
145	K	B	18.99	39.77	8.64	
154	R	B	27.56	52.86	19.44	
155	L	B	0.85	9.75	54.63	Hotspot

Table 3. Consensus of HotPoint prediction of hot spot residues on Skp2 based on 2AST structure. Two pockets were selected based on the two groups of the hotspots.

Trp97, Leu100, Ile108, and Val120 were predicted by all three methods to be hot spots, and Trp127 was predicted to be a hot spot by KFC2B and HotPoint. The first region (referred to henceforth as pocket 1) is close to the N-terminus of Skp2 and is actually within the F-box motif, including Trp97, Phe109, Glu116, Lys119, and

Trp127. The second region (referred to henceforth as pocket 2) has proximity to the corresponding C-terminus of Skp1, formed by the first and second LRR as well as some residues from the F-box, essentially around the variable interface of the F-box region of Skp2 (**Figure 11**).

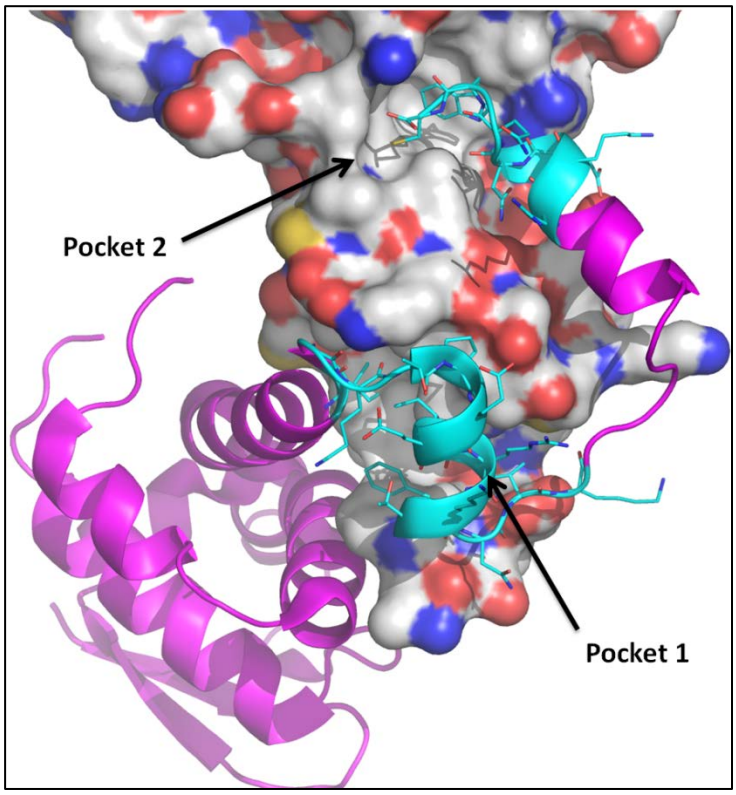


Figure 11. Identified binding pockets on the Skp12/Skp2 complex. Skp2 (molecular surface: gray for carbon, blue for nitrogen, and red for oxygen atoms). Skp1 (purple ribbon). Cyan (Skp1 residues interacting with Skp2).

2.2.2 Hot Spots on Cks1/Skp2/p27 interface

After the consensus analysis of the KFCA, KFCB, PCRPi, and HotPoint programs, the residues determined to be the most significant for binding were found to center around the PhosphoThreonine187 (TPO187) and Glu185 residues of the p27 peptide (**Figure 12**). We selected the residues for which to do virtual screening (using the GOLD software) based on these two regions (calling the region around Glu185 “Pocket1” and the region around T187 “Pocket2”) and, applying techniques similar to

those in de novo drug design, we biased our results towards hits that showed similar properties and positions to that of the residues in those areas.

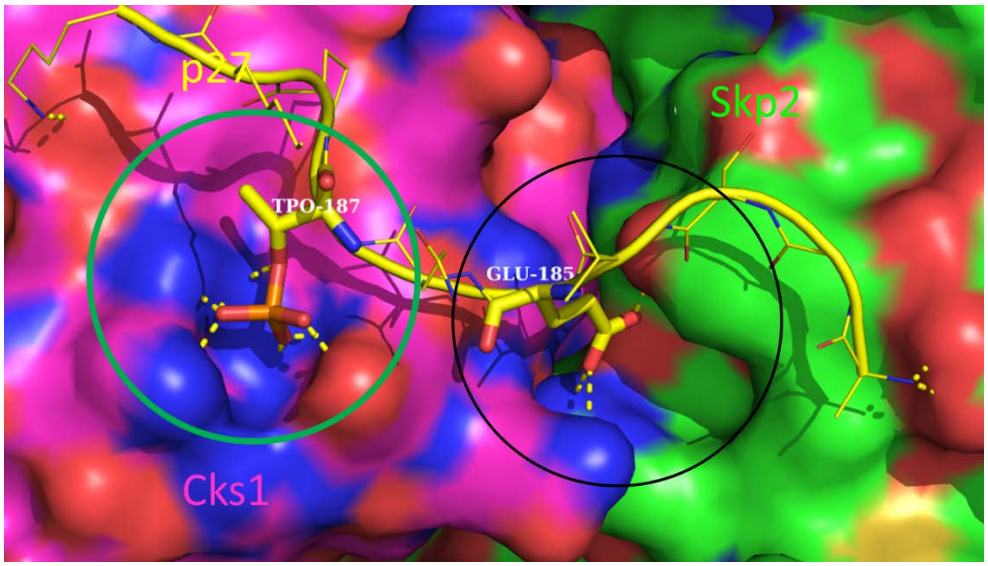


Figure 12. Cks1/Skp2/p27 binding site. TPO187 and Glu185 are both significant for binding. Virtual screens were centered on these two points separately, and one screen included both points together.

The residues that were selected on Cks1 were: Tyr8, Ser9, Lys11, Arg20, Glu40, Arg44, Gln49, Gln50, Ser51, Gln52, Trp54, and Arg71. For Skp2, the residues were: Arg294, Asp319, Arg344, and Tyr346 (**Table 4.** Consensus of 4 hot spot predictions on the Skp2/Cks1/p27 interaction site. Residues selected as hot spots highlighted in orange or red.) **(Figure 13**).

Essentially, this was a more straight forward hot spot prediction, as the p27 surface area residues that are contacting Skp2 and Cks1 are much smaller and less intricate than the Skp1/Skp2 interaction site. Most of these residues closely resembled the ones that are in contact with p27 peptide, and, as was mentioned previously, it is known that mutations on the Thr187 and Glu185 residues abolish or significantly impact the ubiquitination of p27(242). Residues with at least two programs reaching a consensus were considered “mild” hot spots (highlighted in orange), while more than

two consensus predictions were considered “strong” hot spots (highlighted in red). The number of polar contacts were also calculated by MOE and PyMOL, and these also had good agreement with the PCRPi and KFC2 models. HotPoint was perhaps overly sensitive in this prediction space with many potential false positives, but it had no false negatives.

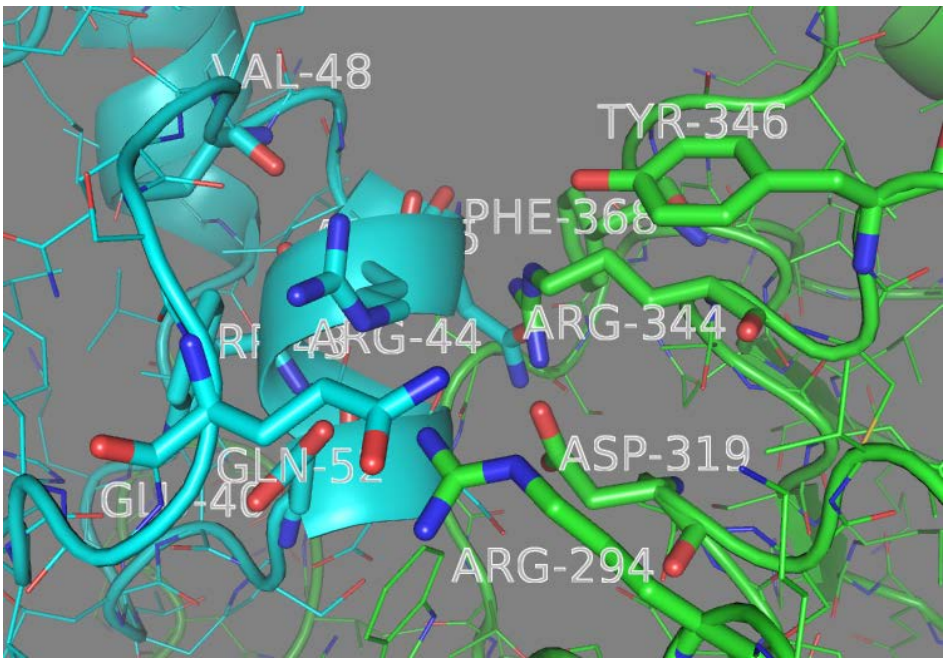


Figure 13. Hot spot residues for Cks1 (blue)/Skp2 (green) interface.

Chain	Res#	Res	Hotpoint	KFC2A	KFC2B	PCRPi Rank	Polar Cntacts
B	265	Trp	H	H	H	3	1
B	291	Ser	H				1
B	294	Arg	H			8	1
B	318	Ser	H			9	
B	319	Asp	H			7	1
B	343	Ser	H				
B	344	Arg	H	H	H	1	4
B	368	Phe	H			13	
B	393	Trp		H	H	2	
B	398	Arg	H		H	4	3
C	35	Thr		H		9	
C	36	His			H		

C	37	Leu	H	H	H	10	2
C	38	Met	H				
C	40	Glu	H			2	
C	41	Ser	H	H		1	3
C	42	Glu	H	H	H	3	3
C	44	Arg	H			5	
C	45	Asn	H	H		6	

Table 4. Consensus of 4 hot spot predictions on the Skp2/Cks1/p27 interaction site. Residues selected as hot spots highlighted in orange or red.

In conclusion, HotPoint and KFC models appeared to have the best predictive capacity for the two interaction sites on Skp2, but the use of a larger consensus (programs such as FTMap and PCRPi), as well as a visual and robust structural analysis was needed to increase the confidence in the predictions. HotPoint appear to perform better for larger and more elaborate surface area interactions (the Skp1/Skp2 interaction), and its potential bias towards larger residues is a means by which it can have this strength, but it hinders for smaller interactions. Other studies have shown it to be highly sensitive when compared to other methods (254). FTMap and KFC (and to a lesser extent, PCRPi) showed their strength in the smaller sites, but performed poorly on the larger complexes. The performance of KFC on smaller sites is likely due to their use of a training set that primarily involves hormones, inhibitors and small fragment proteins, and not many large complexes (84).

These predictions formed the basis for the sites (the general area as well as which specific residues to focus on) on which to perform the virtual screening. The centroid of the screen was selected such that the area encapsulated by the docking program (GOLD uses a radius, HiPCDock uses a cubic unit) contained all the residues selected as hot spots.

2.3 Results: Evaluation of Skp2 Virtual Screening

2.3.1 Ligands Submitted for Testing for Skp1/Skp2 interface

For the Skp1/Skp2 virtual screen, 41 total ligands were chosen as compounds to be tested for pocket 1 with 16 compounds listed as high priority. For pocket 2, 16 total compounds were selected, and 11 of these were high priority. Again, we assigned high priority to the compounds in which we had the highest confidence in these being active biologically based on their pose and diversity in the virtual screen. Only the high priority compounds were ordered for testing in the thought that if none of the high priority compounds were active, the medium and low priority compounds would be selected for a second round of testing.

HiPCDock Results: The HiPCDock docking scores of the 50,080 compound DiverSet library into pocket 1 and pocket 2 are summarized below (**Table 5**). The histograms of the energies are close to extreme value distribution, and are displayed below (**Figure 14**), (**Figure 15**).

	Pocket 1	Pocket 2
Mean	-5.43	-4.12
Median	-5.39	-4.14
Mode	-5.55	-4.23
Maximum	-0.83	-0.00026
Minimum	-12.31	-10.95
Standard Deviation	1.138	1.18

Table 5. HiPCDock docking sores of DiverSet into Pocket 1 and Pocket 2

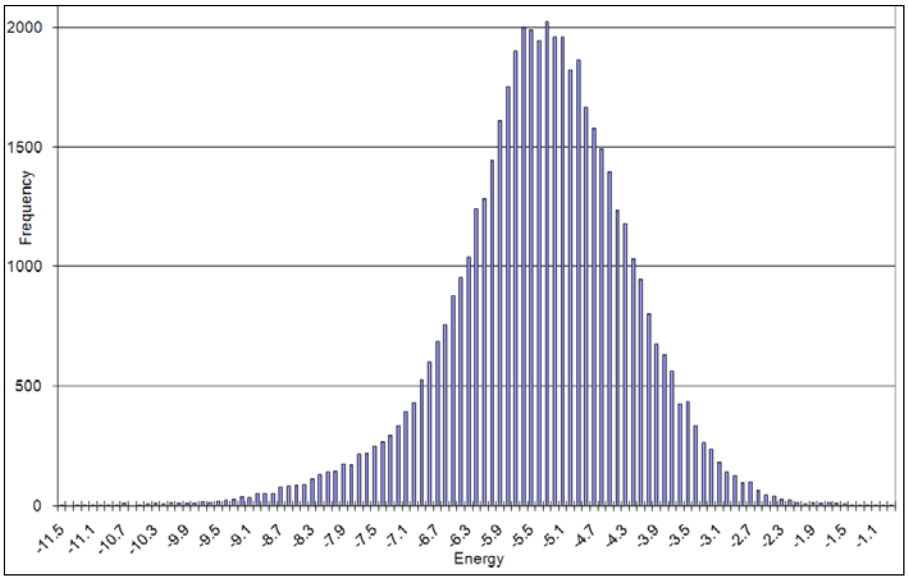


Figure 14. Histogram of energies of Pocket 1 HiPCDock docking results.

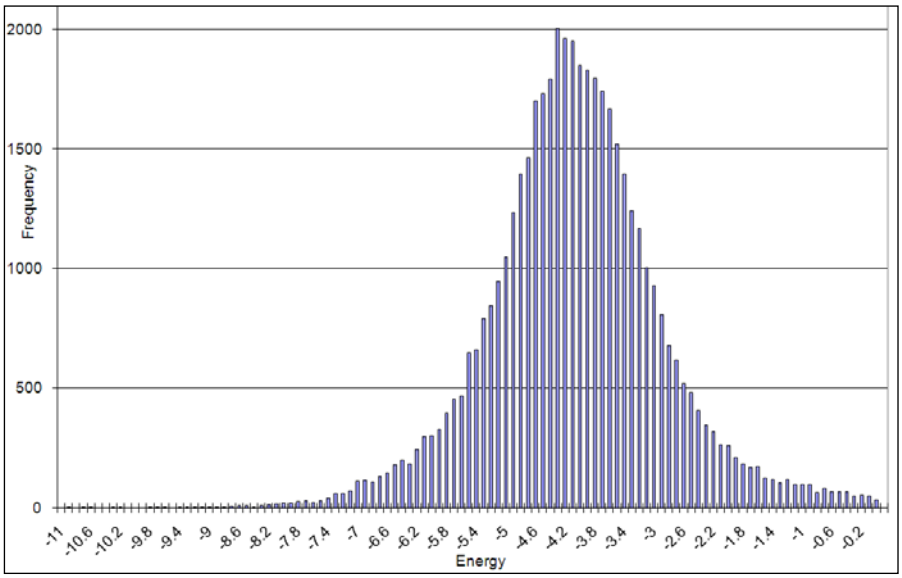


Figure 15. Histogram of energies of Pocket 2 HiPCDock docking results.

Gold Results

In GOLD, the default scoring function is the GoldScore; this is comprised of four components from the equation:

$$f = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_\mathrm{int}} + S_{vdw\_\mathrm{int}}$$

Where S_{hb_ext} is the protein-ligand hydrogen bonding score and S_{hb_int} is the internal hydrogen bonding of the ligand. S_{vdw_ext} and S_{vdw_int} are the scores arising from weak

external and internal Van der Waals forces, respectively. The top Gold fitness scores for the 4,000 compounds (2000 for pocket 1, 2,000 for pocket 2) are summarized below in table form (**Table 6**), and histograms (**Figure 16**), (**Figure 17**).

	Pocket 1	Pocket 2
Mean	37.76	35.65
Median	37.93	35.91
Mode	42.01	36.84
Maximum	62.55	56.15
Minimum	-44.99	8.29
Standard Deviation	8.59	5.54

Table 6. Gold docking sores of DiverSet into Pocket 1 and Pocket 2

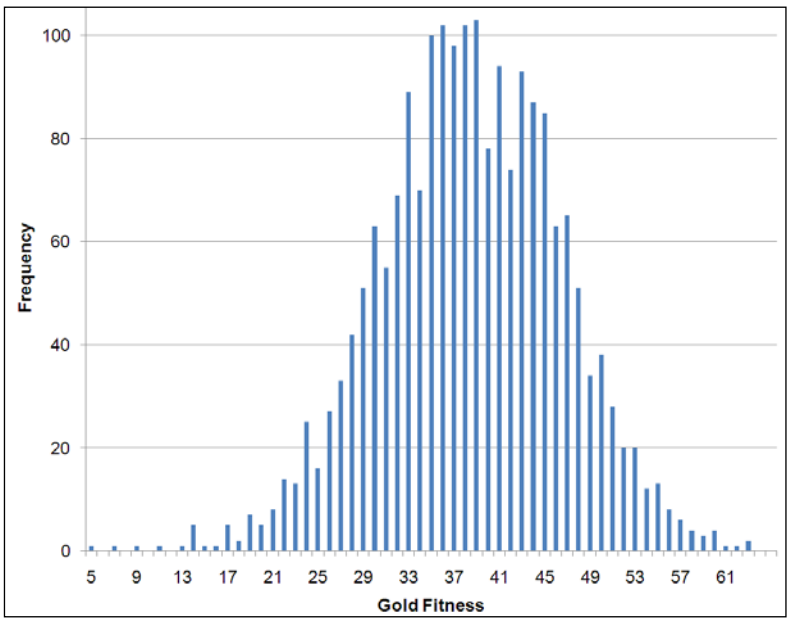


Figure 16. Histogram of energies of Pocket 1 Gold docking.

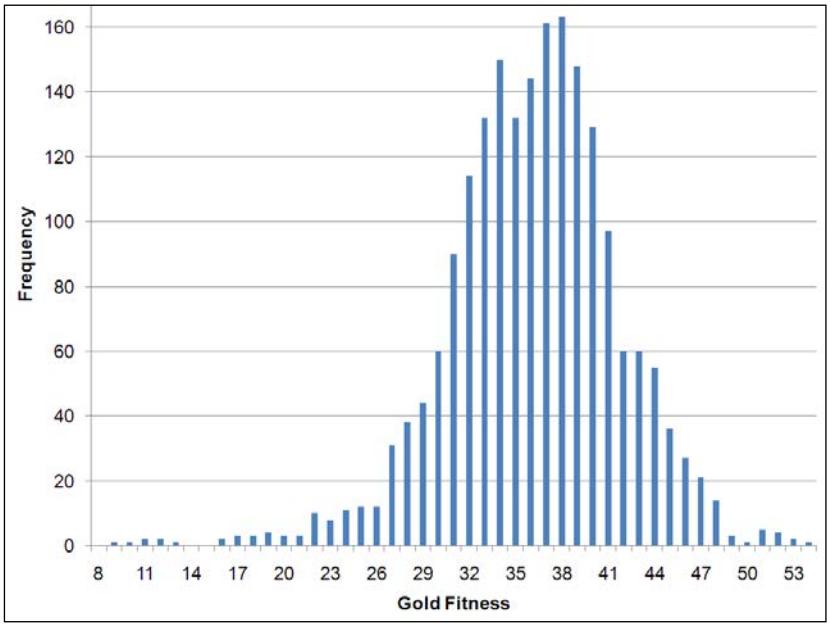
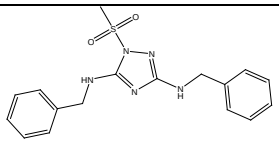
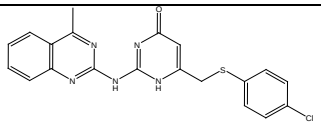
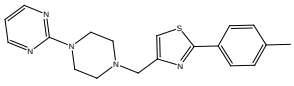
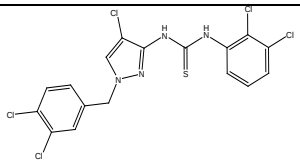
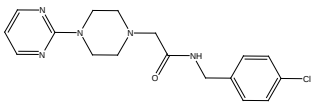
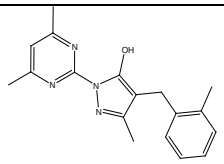
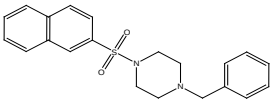
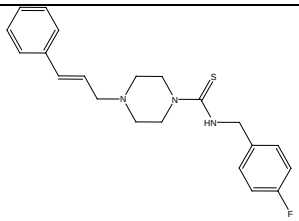
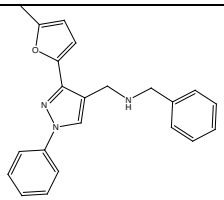
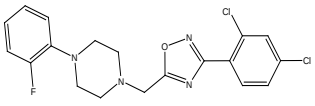


Figure 17. Histogram of energies of Pocket 2 Gold docking.

The biological testing of the Pocket 2 compounds revealed almost all compounds were poor binders except for one: 9040817. However, investigation of the mechanism revealed it was not acting in the mechanism we proposed. Also, secondary screens were showing that 9040817 was not effective (**Table 7.** Top 16 Compounds selected for biological testing for Pocket 2. Ligands marked as “High” priority were tested biologically. Most of these compounds were not active except for 9040817.

).

#	Structure	ID	Gold Fitness	Auto Dock Score	Class	Priority	In vitro binding
1		7974443	56.15	-6.47	1	High	>100 µM

2		7694040	52.21	-8.35	2	High	>100 µM
3		9043923	51.75	-7.05	3	High	>100 µM
4		7839261	50.81	-6.31	4	Med	Not tested
5		9043648	50.12	-6.89	3	Med	Not tested
6		9037036	47.94	-6.5	5	High	>100 µM
7		5570289	47.31	-6.2	6	Med	Not tested
8		7818158	47.28	-6.4	7	High	>100 µM
9	Redacted	Redacted	47.27	-6.68	8	High	5 µM
10		7740910	46.89	-6.7	9	High	>100 µM
11		7739449	46.79	-6.91	3	Med	Not tested

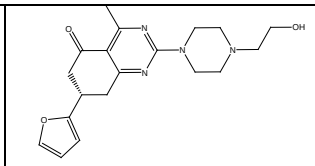
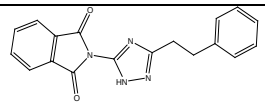
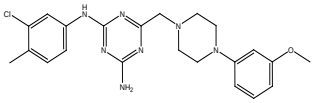
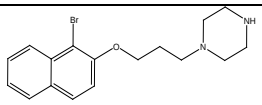
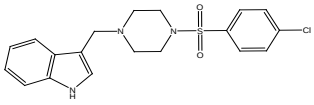
12		9062916	46.43	-6.42	10	High	>100 µM
13		7957666	46.39	-6.51	11	High	100 µM
14		7726220	46.27	-8.56	3	Low	Not tested
15		7631104	46.11	-8.01	12	High	100 µM
16		6037495	45.64	-6.18	7	High	>100 µM

Table 7. Top 16 Compounds selected for biological testing for Pocket 2. Ligands marked as "High" priority were tested biologically. Most of these compounds were not active except for 9040817**.**

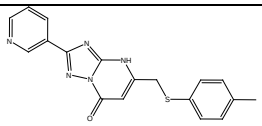
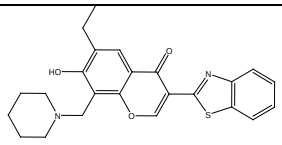
	Structure	ID	Gold Fitness	Auto Dock Score	Class	Priority	In Vitro Binding	Cell Sensitivity (IC_{50})
8		9040817	57.31	-8.19	3	High	5 µM	>20 µM
19	Redacted	Redacted	54.47	-8.04	10	High	5 µM	5 µM
35	Redacted	Redacted	51.46	-8.52	14	High	5 µM	5-10 µM
39	Redacted	Redacted	51.17	-8.03	17	High	5 µM	10-20 µM
41		5572358 (ZL-25)	50.71	-7.95	18	High	5 µM	5-10 µM

Table 8. Results of cell sensitivity (secondary screen) studies. Ligands #35 (which has its structure redacted) and #41 (5572358) were found to be the most sensitive and fit the mechanism of action.

The top performing compounds from both pocket 1 and pocket 2 were submitted to cell sensitivity studies. The Cell sensitivity studies showed that compounds Redacted and 5851912 did not fit the mechanism proposed. These likely were achieving the sensitivity values due to another, off-target mechanism of action (not via targeting Skp2 and p27 ubiquitination) and were not pursued further. The compounds that were found to be most active and fit the mechanism were 5572358 (which became the lead, ZL-25), and a second compound (which was also known internally as ZL-22). ZL-22 was only slightly less active than ZL-25, so ZL-25 became our primary lead.

2.3.2 Ligands Submitted for Testing for Cks1/Skp2/p27 interface

In regards to the Cks1/Skp2/p27 site, the unique features of this particular target required a more unconventional approach in regard to the virtual screening methods. The crystal structure shows that Cks1 binds to the concave leucine-rich repeat (LRR) domain on Skp2 (the interface buries ~ 1850Å^2 of solvent accessible surface), and p27 forms contacts with both Cks1 and Skp2, with interface sizes of ~740Å^2 and ~400Å^2, respectively. These interfaces form what is essentially an atypical, three-part protein complex on which we focused our inhibitor design efforts. The use of three screens covering the two pockets individually and then both taken simultaneously was used to maximize the possibility of discovering a ligand large enough that could hit both pockets together, (such that might have a stronger potential

to block p27 binding) but still retaining the potential of finding a ligand that had a high affinity for only one pocket. The top hits from the two smaller targeted screens were merged into only 5 compounds with high priority that were tested (**Table 9**).

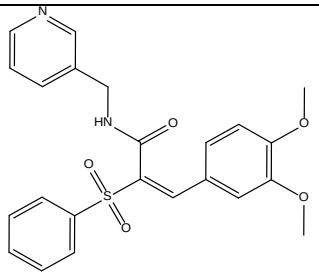
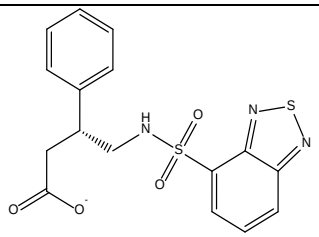
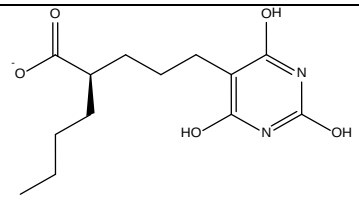
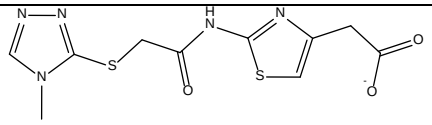
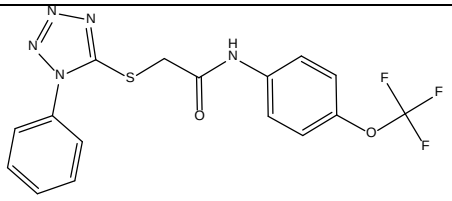
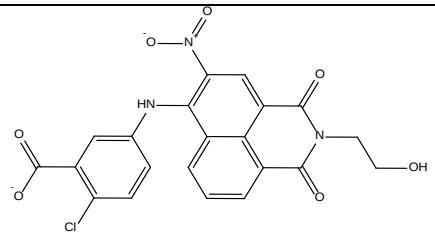
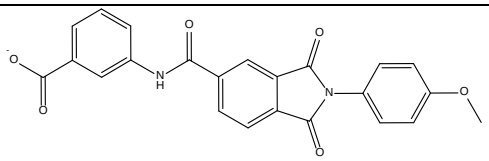
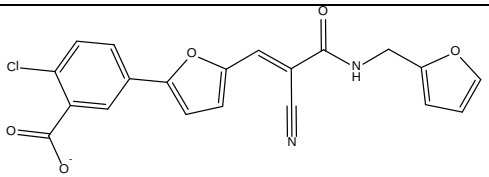
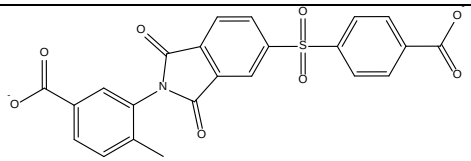
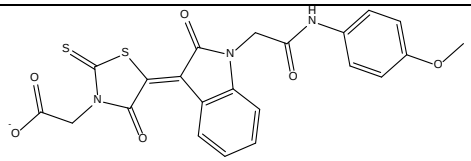
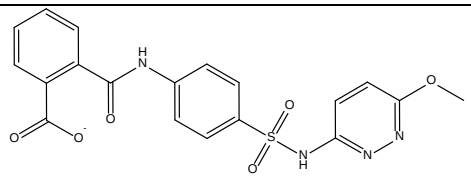
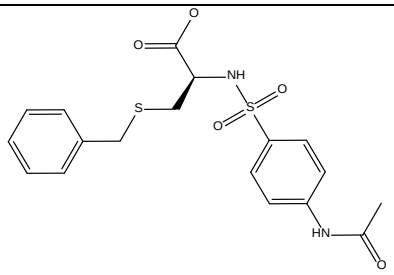
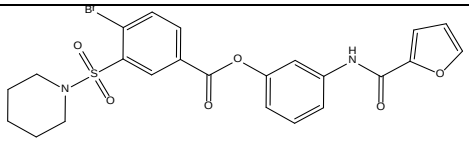
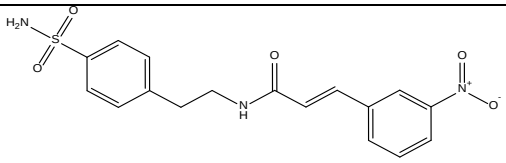
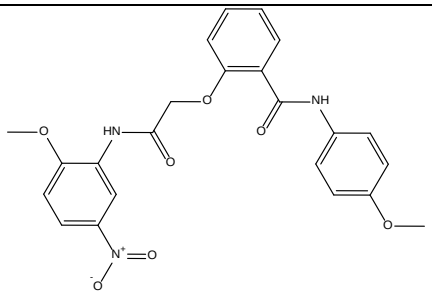
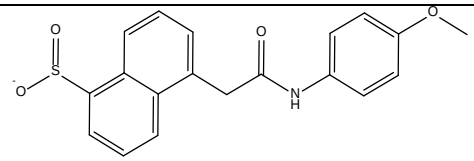
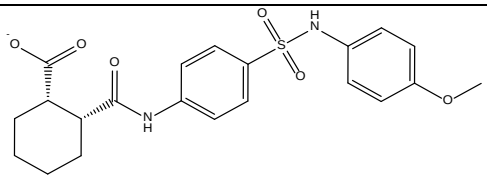
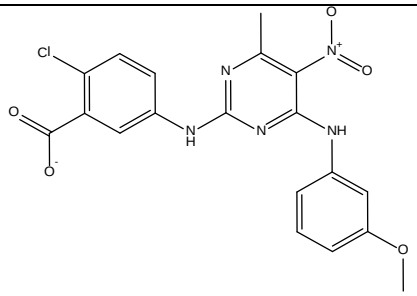
#	Structure	ID	Gold Score	MolWt	logP
1		ST029405	65.9258	438.50	3.033
2		ST041756	65.8194	376.43	2.729
3		R732486	62.0391	283.30	2.333
4		R549606	61.2877	312.35	0.534
5		ST027355	60.3729	395.36	3.451

Table 9. Top hits from two small targeted screens with high priority that were biologically tested. #1 (ST029405, referred to later as ZM-405) was found to be active and fit the mechanism.

From the wide screen that incorporated both pockets, there were 14 compounds selected, five were considered high priority. Due to the low number of compounds listed here, all were ordered and all were tested (**Table 10**).

#	Structure	ID	Gold Score	MolWt	logP
1		ST016939	75.95	454.80	3.372
2		ST031312	74.82	415.38	3.506
3		ST024634	73.99	395.77	2.737
4		ST031378	73.44	463.42	3.543
5		ST040033	73.057	482.51	1.92
6		ST011908	72.54	427.41	3.176

7		R508276	72.09	407.48	2.381
8		ST004338	72.06	533.39	3.759
9		ST031897	71.68	375.40	2.125
10		ST016921	71.29	451.43	3.476
11		ST024091	71.26	354.40	2.375
12		ST040186	71.24	431.48	2.708
13		ST021173	71.19	428.81	3.674

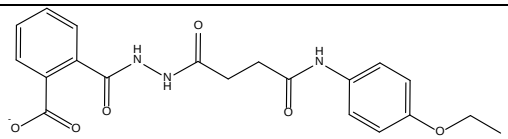
14		L440442	70.76	398.39	1.78

Table 10. Selected compounds for testing from broad based screen covering both T187 and Glu185 pockets. #14 (L440442, later known as ZM-442) was originally considered active in initial screens but was not confirmed in secondary assays.

2.3.4 Discussion of findings from Virtual Screens

The 2AST crystal structure was chosen for its high resolution and its inclusion of the p27 and Cks1 peptides. Consideration of the residues that contacted Skp2 from Skp1 was a crucial step in understanding which sites would lead to the disruption of this complex, and the site finder in MOE helped to validate pocket 2 as a reasonable location to perform one of the virtual screens. The choice of the DiverSet library as the primary screening ligand set ensured that adequate chemical diversity was present in the virtual screening. The diversity stems partially from the more than 260,000 unique 3-point pharmacophores present within DiverSet. It has high diversity in drug-like compounds, and applied filters remove unsuitable compounds such as: undesirable chemical groups (e.g. Michael acceptors, crown-ethers & analogs, disulfides, epoxides), structural, salt and tautomeric duplicates (241).

Using HiPCDock as the first High-Throughput Docking (HTD) tool for screening because of its, speed, efficiency and ease of use in a High-Performance Computing, multiple-CPU cluster environment. Its basis on the AutoDock framework allows for easy automation and high levels of control over the parameters involved, and its statistical-based bioinformatics model towards structure-based HTD gives it an

advantage over other docking tools. The binding free energy distribution is very close to extreme value distribution, which allows us to calculate the probability and expectation of the binding free energy of the ligand-receptor complex; from these calculations, a statistical significance of the prediction can be evaluated.

By docking the top results from HiPCDock into GOLD, we were able to utilize a consensus method to enhance the probability of finding active compounds. GOLD has been shown in multiple studies to have high accuracy in the prediction of binding modes of ligands and it is a more than acceptable program for its use in various techniques when compared with other docking or virtual screening programs (255-257). The clustering of the GOLD results based on MACCS fingerprints increases the relative diversity of the pool of molecules selected for biological testing and removes structurally similar compounds.

The top scoring 130 compounds cluster centers were visualized individually by hand in PyMOL to maximize the potential of finding ligands that would prove to be active biologically. This higher scrutiny at this stage not only ensured that there was more than adequate diversity in the final compounds; it confirmed that the compounds resided in the desired interaction site and therefore had a high potential of inhibiting the interaction with Skp1. Also of note, our analysis was independent of the two docking scores that we obtained; this made for our analysis to be more qualitative than quantitative.

Skp2, when in complex with the different bound forms of each ligand, gives us insight as to what features lead to this good activity. The ligand 5572358 (the 41st

entry in the above table, referred to as ZL-25 in the remainder of this study), when docked to Skp2, is indicative of the structures of other ligands that showed high docking scores, and good biological activity. Also of note is the pi stacking occurring towards Trp97, with an average distance of 3.27Å to this residue. The ligand ZL-22, which also shows good biological activity, also shows a single hydrogen bond, but occupies a lipophilic region bound by Ile108, Trp127, Phe109, and Leu105. Again, the pi stacking appears to be strong between the ZL-22 ligand and the Trp97 residue, with an average distance of 3.1Å.

2.3.4 Summary of findings from Virtual Screens

In total, from the initial virtual screens there were two compounds initially considered leads for both Skp2 regions. ZL-25 and ZL-22 for the Skp1/Skp2 interface (**Figure 19**), and ZM-405 with ZM-442 for the Cks1/Skp2/p27 project (**Figure 18**). From these, ZL-25 and ZM-405 were considered the primary leads on which lead optimization would be performed (**Figure 20**).

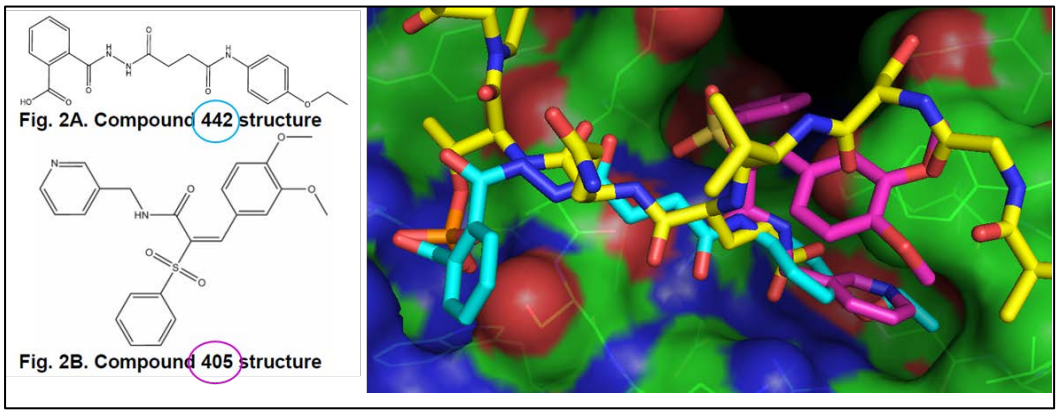


Figure 18. Top two p27 inhibitors discovered via virtual screen (left) and proposed binding poses (right). Compound ZM-442 (blue sticks) covers both pockets (Thr187 and Glu185), whereas compound ZM-405 (purple sticks) is predicted to bind to just

the Glu185 pocket. Both compounds are capable to block p27 binding (shown in yellow sticks).

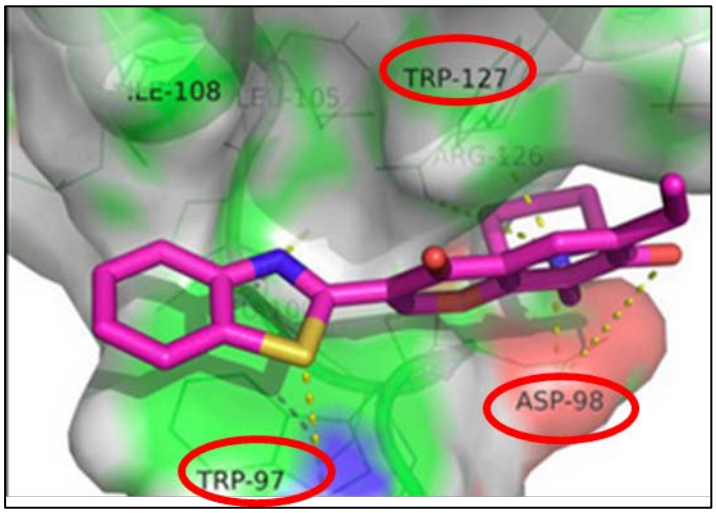


Figure 19. ZL-25 in proposed docked position on Skp2. Significant residues for binding circled.

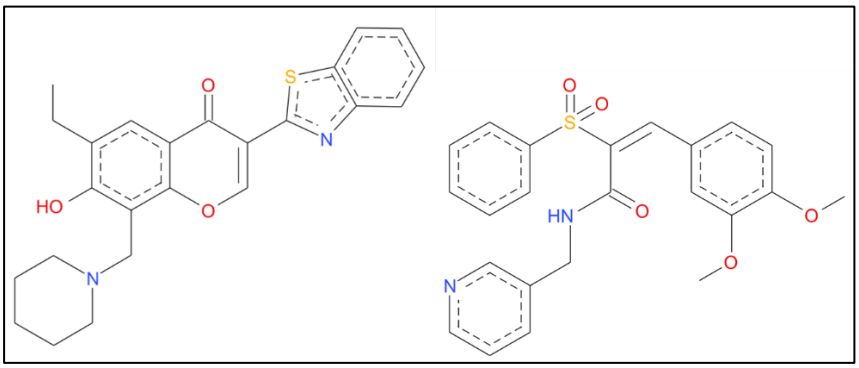


Figure 20. Structures of the two Skp2 inhibitor leads: ZL-25 (left), ZM-405 (right).

2.4 Results: Biological Effects of Virtual Screening Results

2.4.1 In Vitro and In Vivo Data of Skp1/Skp2 inhibitor ZL-25

Portions of this data used with permission from the journal Cell, in the article: *"Pharmacological Inactivation of Skp2 SCF Ubiquitin Ligase Restricts Cancer Stem Cell Traits and Cancer Progression" Cell. 2013 Aug 1;154(3):556-68. (176)*

Our collaborators performed the biological testing of the suggested compounds, the effects of the compounds are presented here. The initial *in vitro* binding assay/screen that identified ZL-22 and ZL-25 was a simple GST-Skp1 pull-down assay at 5μM showing that it was able to completely prevent Skp1-Skp2 interactions (**Figure 21A.**) and do so in a dose dependent manner in PC3 cells (**Figure 21B**). Next, our collaborators tested the ability of ZL-25 to reduce the levels of ubiquitinated p27 in vitro with a Skp2-mediated ubiquitination assay, and again it was able to achieve these effects in the sub-10μM range resulting in their stabilization (**Figure 22)**.

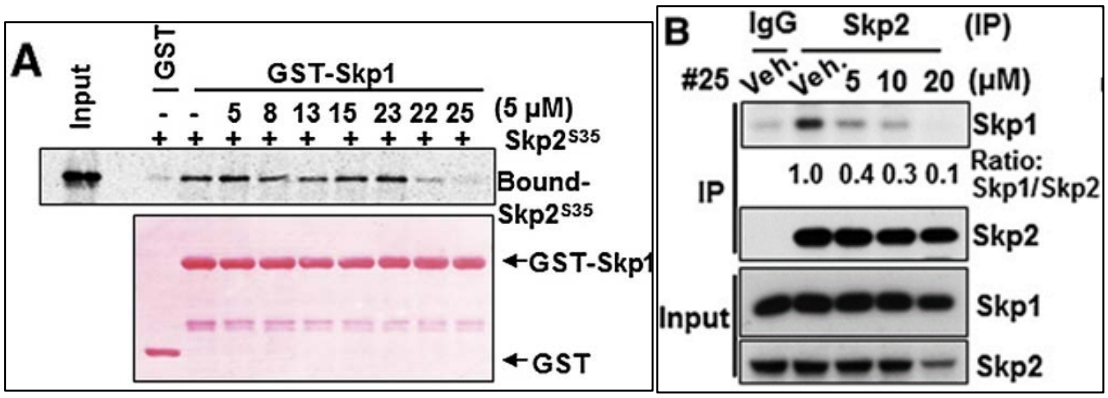


Figure 21. A. In vitro binding assay of Skp1-Skp2 with our without ZL-25 (far right lane). **B.** In vivo dose dependent Skp1-Skp2 binding assay in PC3 cells.

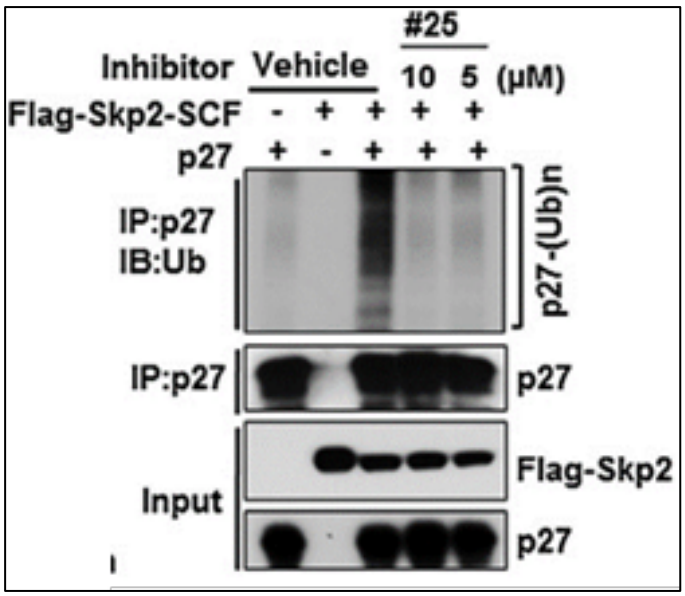


Figure 22. ZL-25 is able to inhibit Skp2-mediated p27 ubiquitination in vitro.

Additionally, consistent with the ubiquitination assay of p27, a dose-dependent induction in the fold change of p21 (another substrate of Skp2) and p27 was seen in PC3 (prostate cancer) cells is apparent with ZL-25. It is not readily apparent as to why the 1μM dose seems to result in a small decrease in the fold change of p21 and p27 (**Figure 23**).

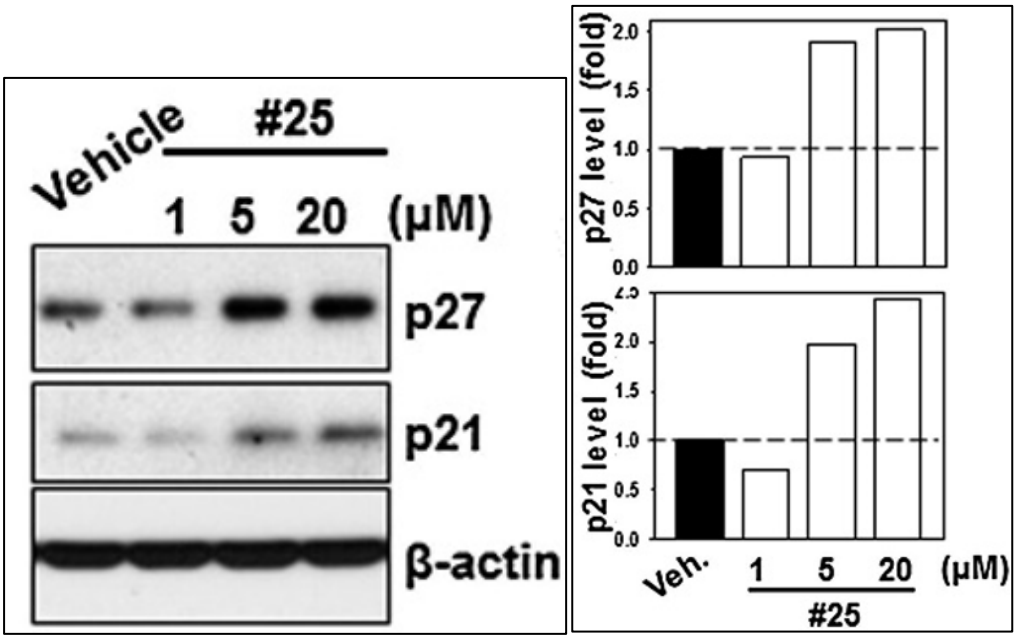


Figure 23. Induction of p21 and p27 via ZL-25. PC3 cells were treated with DMSO or ZL-25 at different doses for 24hrs and harvested for immunoblotting (IB) assay.

As Akt is a substrate of Skp2 for K63-linked ubiquitination and Skp2 is responsible for its activation via this mechanism, ZL-25 was also shown to inhibit Skp2-mediated ubiquitination both in vivo (293T cells) and in vitro (**Figure 24**).

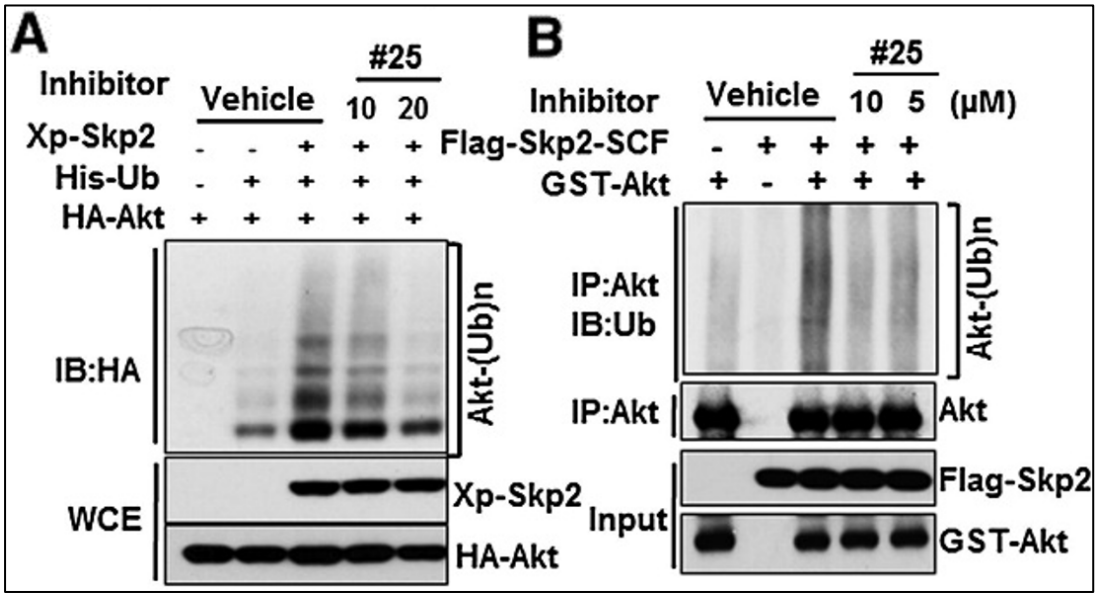


Figure 24. ZL-25 also affects other Skp2 substrates, such as Akt. **A.** In vivo Akt ubiquitination assay showing dose dependent inhibition. **B.** In vitro Skp2-mediated Akt ubiquitination assay.

2.4.2 Confirmation of Binding Mechanism of ZL-25

The predicted binding mode of ZL-25 shows several primary interactions with the F-box domain on Skp2. Most notable are the Trp97, Asp98, Arg126, and Trp127 residues (**Figure 25**). Other nearby residues that contribute to Skp1 binding but not ZL-25 binding include Pro101, Leu105, Ile108, and Phe109.

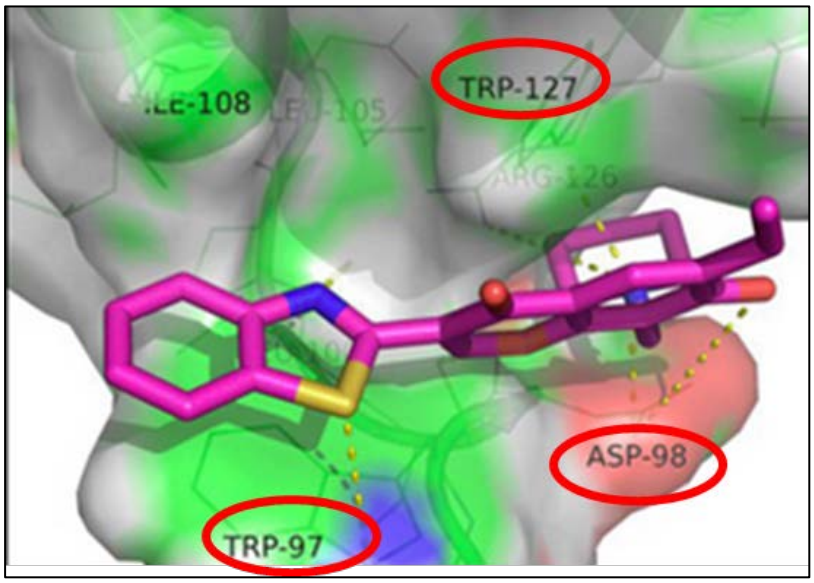


Figure 25. Proposed docking mode of ZL-25 to Skp2 F-box domain. Highlighted residues contribute significantly for ZL-25 binding.

To determine the residues responsible for binding to both Skp1 and to ZL-25, site-directed mutagenesis was carried out on all the above residues. The mutant proteins had their residues mutated to alanine (since it is the least reactive amino acid that does not introduce torsional freedom). Of the mutants Skp2 proteins, only the Skp2 Trp97Ala, Asp98Ala, and Trp127Ala retained their Skp1 binding ability, the others lost their Skp1 binding capacity, but it is worth noting that Ile108 retained 50% of its binding affinity to Skp1 (**Figure 26**). Among the proteins that retained all their Skp1 binding ability the biological activity of Trp97A, Asp98A, and Trp127A were tested for their capacity to bind to ZL-25. It was demonstrated that Trp97A and Asp98A are

resistant to the effects of the inhibitor, which confirms that these residues are responsible for binding ZL-25 to Skp2 and that the predicted binding mode of ZL-25 from the virtual screen is accurate (**Figure 27**).

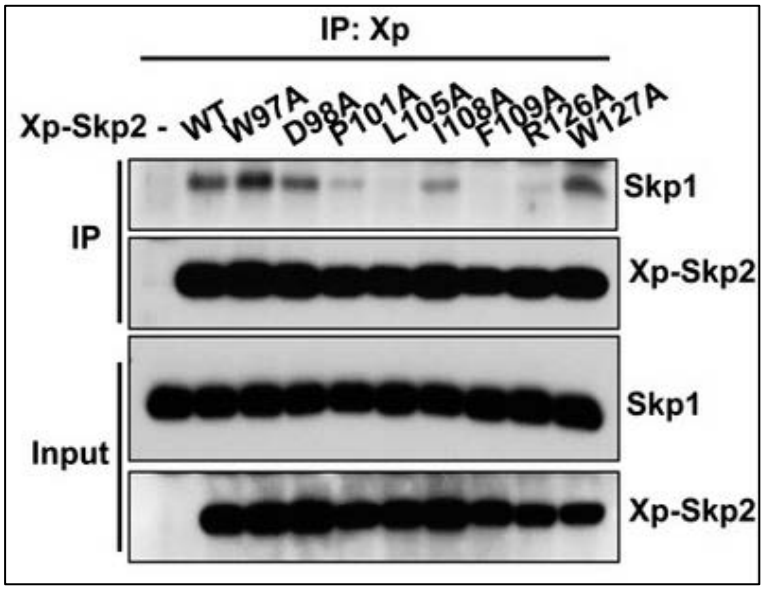


Figure 26. 293T cell Skp1-Skp2 binding assay. W97A, D98A, and W127A all retained their ability to bind to Skp1.

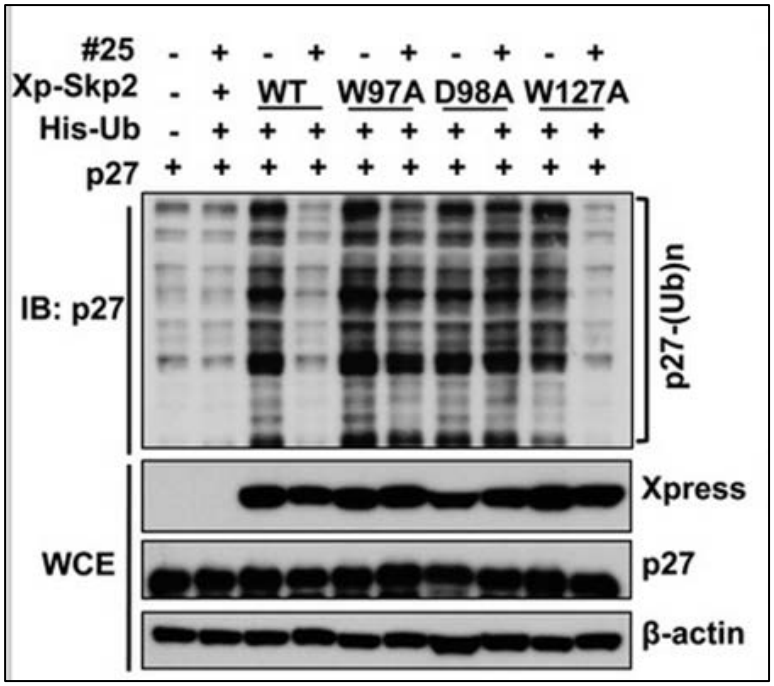


Figure 27. 293T cell ubiquitin binding assay transfected with different mutant constructs and in the presence or absence of ZL-25.

In addition to this mutant data, cell survival was examined of WT Skp2 versus the W97A and D98A mutants. The effect of ZL-25 was examined on survival in two types of prostate cancer cells: Skp2-knockdown and control mutants dosed with ZL-25. The inhibitor had capable effect on the viability of control-knockdown cells in a manner that was dose-dependent, its effects were compromised when compared to the Skp2-silenced cells (**Figure 28**).

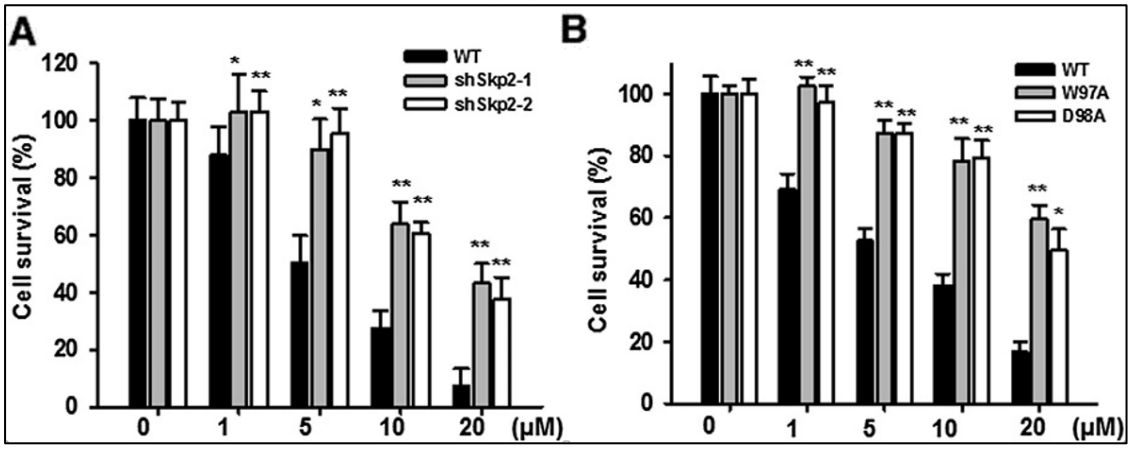


Figure 28. PC3 cells with or without Skp2 knockdown (A) or PC3 cells stably expressed with Skp2 WT, W97A, or D98A mutants (B) treated with various doses of ZL-25, followed by cell survival assay. Cell survival percentage of each stable cell lines treated with various doses of ZL-25 was normalized to that treated with DMSO. Results are presented as mean values ± SD. $*p < 0.05$; $**p < 0.01$.

Next, to determine the specificity of ZL-25, two other highly similar F-box proteins (specifically Fbw7 and β-TrCP) were selected for their potential binding to ZL-25. Despite the drastic effects of ZL-25 on Skp2-mediated ubiquitination of p27 and Akt by Skp2 (shown above), the inhibitor showed no appreciable effects on β-TrCP-mediated ubiquitination of Snail and IκBα (two of the targets of β-TrCP) (258), as well as Fbw7-mediated ubiquitination of c-Jun and MCL-1 (two well described targets of Fbx7) (reviewed in (259)) (**Figure 29**).

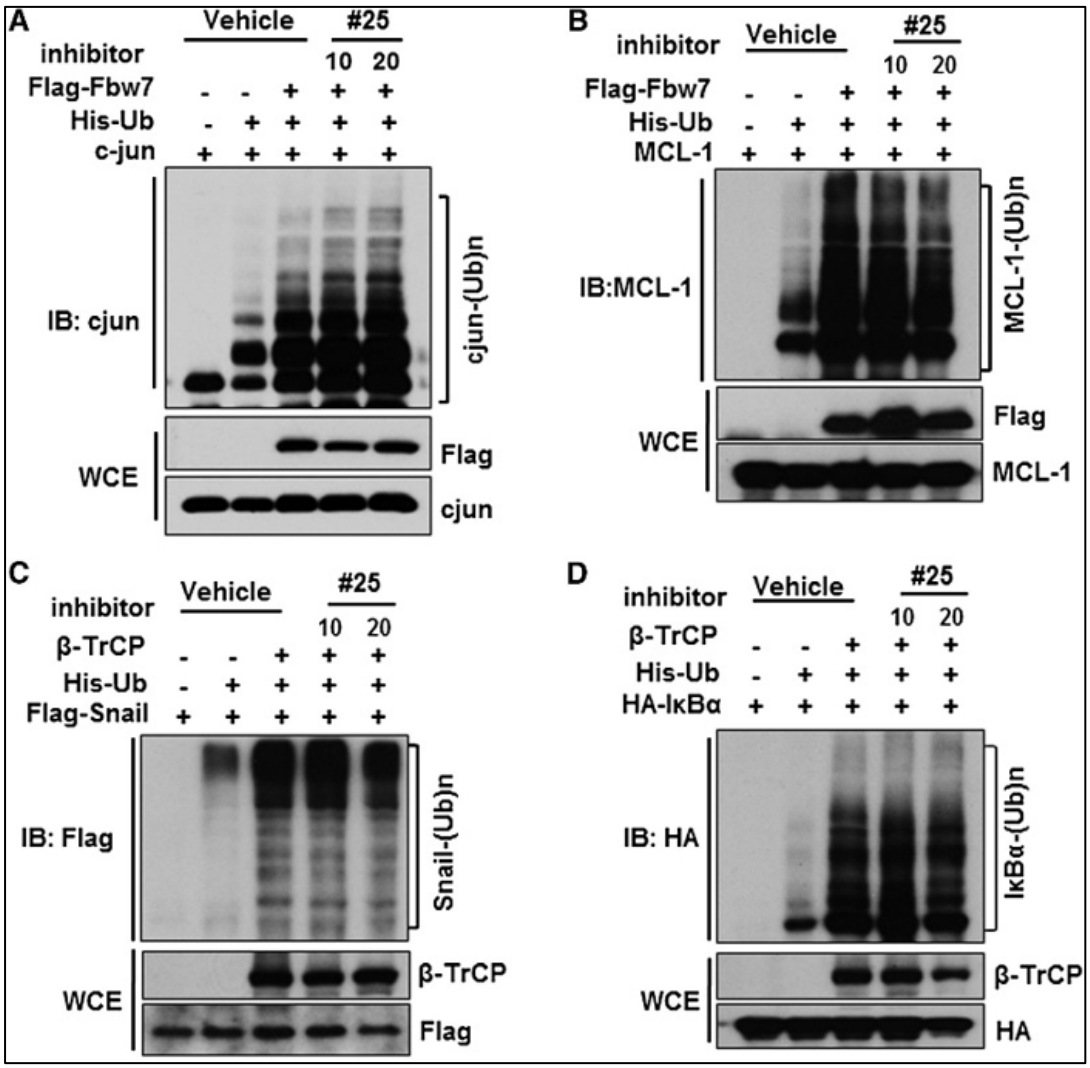


Figure 29. ZL-25 is specific towards diminishing E3 activity of Skp2, but not other F-box complexes such as β-TrCP and Fbw7. 293T cells transfected with various constructs with DMSO or ZL-25 was treated with MG132 for 6hr followed by in vivo ubiquitination assay.

The mutagenesis data shows that ZL-25 binds to the Trp97 and Asp98 residues on Skp2, and this compound is specific for Skp2 and no other highly similar F-box proteins. This prompted a further sequence analysis of the F-box domains of Skp2 compared to other F-box proteins. When the consensus sequence of all 69 human F-box proteins are analyzed, the ZL-25 binding site occurs in a region of the F-box where consensus is low (N = 4,5 in below figure), and Skp2 contains residues not shared by other F-box proteins (**Figure 30**).

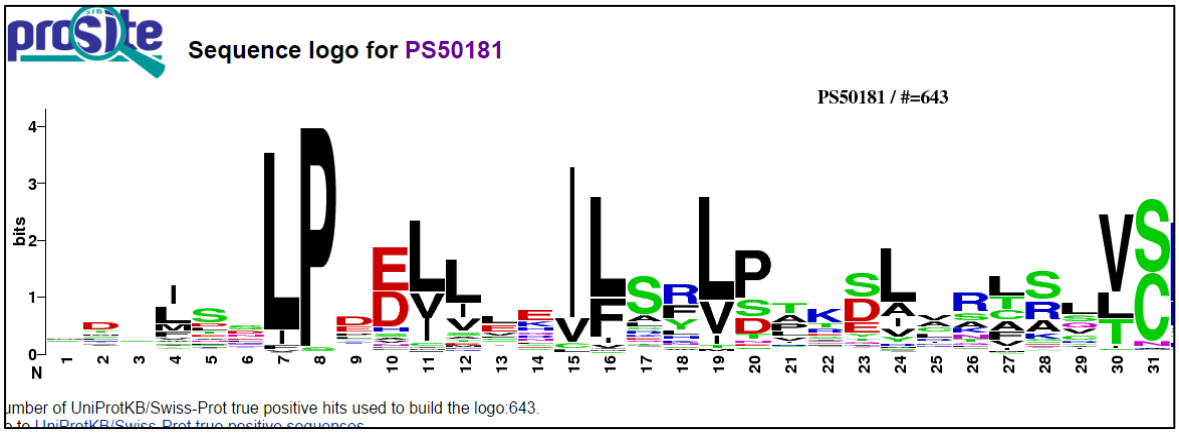


Figure 30. Consensus sequence of first 31 residues of F-box domain shows ZL-25 binding site in region of low commonality to other proteins (residues 4 and 5 in this image, corresponding to Trp97 and Asp98, respectively). The consensus is much higher in the key leucine and proline at residues 7 and 8. Taken from Prosite (260).

There have been other natural products that have been shown to interact with F-box proteins that show similar structures to ZL-25, (which is structurally an isoflavone, which is a natural product based on the flavonoid scaffold) although their specific mechanism of actions and modes/regions of binding have not been confirmed. 1,2,3,4,6-penta-O-galloyl-β-D-glucose (5GG), curcumin, quercetin as well as lycopene all cause downregulation of Skp2 protein in two different breast cancer cell lines, despite the fact that this downregulation did not always correlate with a corresponding p27 upregulation, but when 5GG is used in combination with chrysin, tumor growth was suppressed in xenograft mice (261). The isoflavone wogonin was shown to have anti-cancer effects and was able to down-regulate c-Myc and Skp2 at the protein level and also FBW7 levels were decreased in A549 cells (262). However, as wogonin also has potential to inhibit kinases involved in Ras pathways in melanoma cell lines and has affinity to matrix metalloproteinase-9 (MMP-9) in HCC (263, 264), it is possible there might be an underlying mechanism of which Skp2 is a minor player.

One issue with flavonoids as treatments is that they are readily and easily metabolized by phase II enzymes such as glucuronosyltransferases (UGTs) and sulfotransferases (SULTs), reducing their half-life drastically. Isoflavones such as genistein (ZL-25 is an isoflavone) however have been shown to not be metabolized by SULT1A3 (the major SULT in humans), but by SULT1A1, which is expressed in much lower amounts (265). Genistein, when compared to other isoflavone analogs, has also been shown that the formation of the genistein metabolite did not have comparable excretion and formation rates, showing that the intestinal disposition of isoflavone compounds was structurally dependent (266).

2.4.3 In Vitro and In Vivo data of Cks1/Skp2/p27 inhibitor ZM-405

Our collaborators examined the effects of ST029405 (which was renamed to ZM-405) and found it to be the most active compound. As similar to the initial screen of ZL-25 via Co-IP of Skp1 and Skp2, this screen used Skp2 and p27. ZM-405 was able to disrupt p27-Skp2 interactions (specifically, binding to the SCF complex, as Cks1 accessory protein is required to bind to p27) and inhibit the ubiquitination of p27. Immunoprecipitated p27 and Normal Ig (NIg) blotted with anti-Ub antibody show disrupted p27 binding to Skp2 and decreased p27 (**Figure 31**).

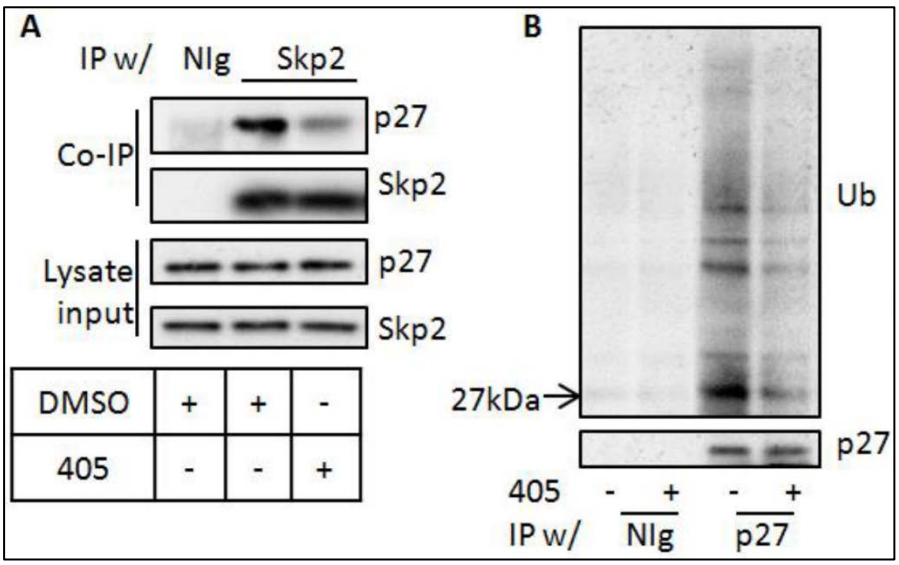


Figure 31. Initial activity discovery of ZM-405. **A**. ZM-405 disrupts p27-Skp2 interactions via Co-IP. **B**. ZM-405 also inhibits p27 ubiquitination. NIg: Normal Ig

To check the specificity of ZM-405, it was tested on a downstream target of Skp2 E3 ubiquitin ligase, the retinoblastoma member p130, which is another target of Skp2 (although not as significant as p27) (132). ZM-405 did not disrupt p130 as much as it did p27 at 10μM doses, indicating it also has good specificity towards its target of p27 (**Figure 32).**

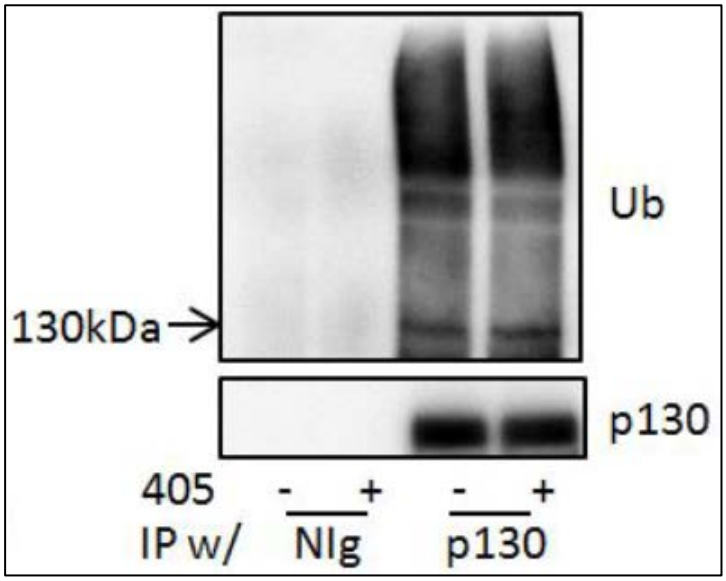


Figure 32. ZM-405 does not target p130 for ubiquitination. At 10μM doses, ZM-405 shows it is specific for p27.

In light of the knowledge that p27 is only effective when in the nucleus and is targeted for ubiquitination by Skp2 primarily in the nucleus (after it is phosphorylated on Thr187 (267) in this study, an attempt was made to look at the endpoint of restore nuclear p27 localization so that Skp2 and the SCF complex will be inhibited in the nucleus, and not the cytoplasm. ZL-405 was able, in a dose-dependent manner to restore nuclear localization of p27 and did not increase cytosolic levels. Again, consistent with the selectivity that was predicted for targeting the Cks1/Skp2 binding site, our collaborators found that it had these effects on restoring nuclear p27 while not having an effect on p130 in the nucleus or in the cytoplasm (**Figure 33**).

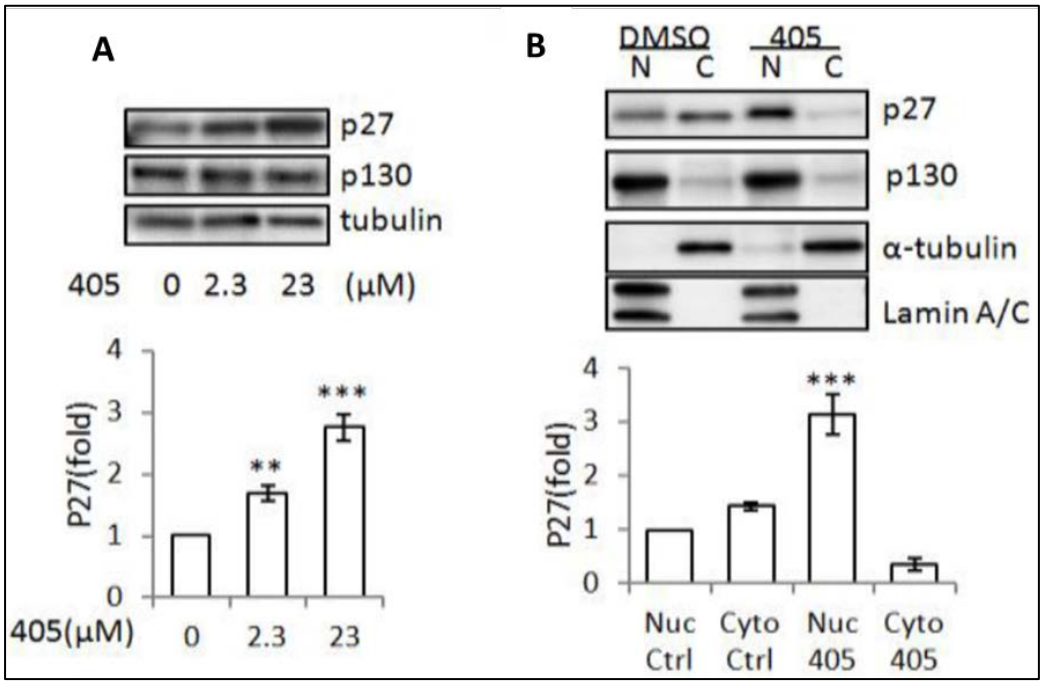


Figure 33. Results of ZM-405 testing on restoring nuclear p27 levels. **A.** Western analysis shows selective p27 (but not other targets such as p130) levels increase. **B.** Total p27 increase is associated with increased nuclear and decreased cytoplasmic.

Another structural analysis of the Cks1/Skp2 interaction site in comparison to other F-box proteins shows that the region of Cks1/p27 binding does not share homology with other proteins, indicating that the specificity of Cks1/p27 binding

might be due to the differing sequence and explains why p130 is not affected in inhibitor treatment (**Figure 34**).

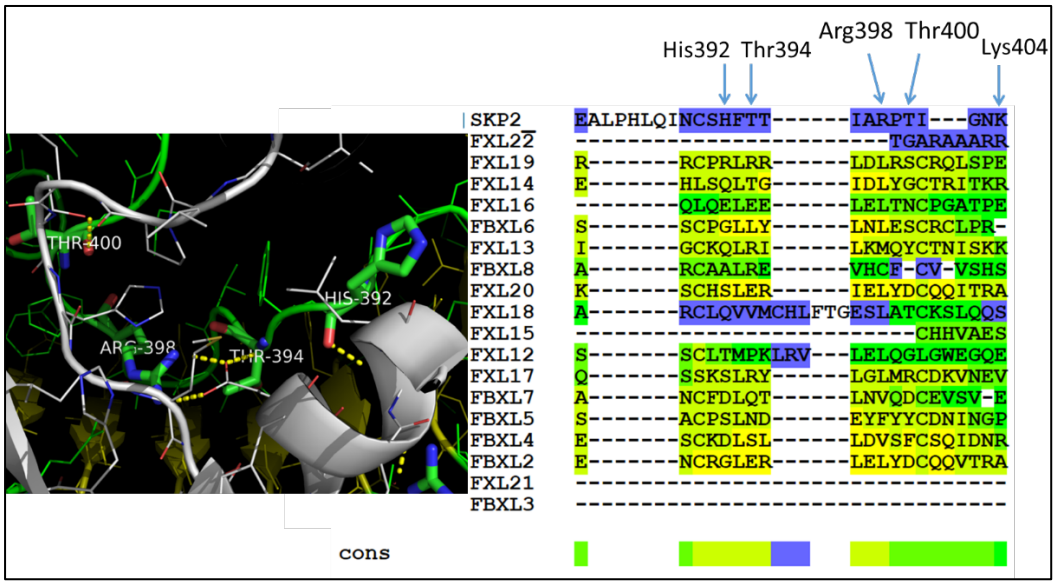


Figure 34. Conservation of residues of seat-belt region of Skp2 (top row) compared to other F-box proteins. Warmer colors indicate higher conservation.

2.5 Summary of Results and key Conclusions from Hot Spot study and Virtual Screening Data

After a rigorous study of the structure of Skp2 on two key regions (the Skp1/Skp2 interface and the Cks1/Skp2/p27 interface), an ensemble of computational hot spot prediction techniques were used on these two regions and the key residues that were suggested to be hot spots became the basis on which to perform multiple *in silico* high-throughput virtual screens using multiple large datasets of drug-like and lead-like compound libraries. Two lead compounds were discovered and tested in various biological systems and found to have good initial activity and good specificity for different areas of Skp2 (ZL-25 and ZM-405). In the next chapter, multiple

techniques will be used to analyze their binding and study their effects upon binding in order to suggest modifications to increase their potency and specificity.

Chapter 3: Binding Analysis and Pre-Clinical Development of Inhibitors

3.1 Introduction

From the hot spot analysis of Skp2 (both in the Fbox/Skp1 region and the Cks1/p27 region, we identified two distinct (but yet distant and with differing effects) regions on which to conduct virtual screens. The subsequent virtual screens gave compounds that showed biological activity in both regions on Skp2.

Here our aims are to build off of the initial biological results obtained from both the Skp1/Skp2 and Cks1/Skp2/p27 inhibitors and develop them for future clinical use. This involves many steps, including: increasing the binding affinity of the inhibitors to their targets, maintain specificity of inhibitors to the targets, and minimize off-target effects, evaluate potential overt toxicity events, and maximize inhibitor's ability to stay in tumor cells and maximize half-life and minimize metabolism. A major question that we also hope to answer is how are the inhibitors binding specifically to their targets, and how are the targets responding to ligand binding. This is especially important and difficult, as all the inhibitors presented here are PPIs, so there will be an interplay of not just two elements in the investigation of the ligand binding.

A concern with the Skp1/Skp2 inhibitor (ZL-25) is that, as was confirmed by the mutagenesis, it binds to Trp97 and Asp98, which are residues that are two away from the end of the Skp2 crystal structure in the 2AST PDB entry (242). The concern here is that the remaining 95 residues on the N-terminus tail of Skp2 very well might be significant for our study of ZL-25 binding, and certainly the N-terminal tail of

Skp2 must have significant interactions with Skp1 as well. Indeed, it was shown, when a mutant Skp2 was created without the 95 residue tail, the *in vitro* binding to Skp1 was greatly enhanced (data not shown), therefore it also must be oriented in such a way that would prevent its binding. Additionally, the nuclear localization signal is a motif in the region of residues 67-73.

Therefore, to address this concern, before conducting binding analysis, a homology model was built that recreated the 95 residue N-terminal tail of Skp2. Homology modeling is a method that is well established and has been documented widely(268), and it is based on the assumption that proteins that possess similar sequences must also share similar three dimensional structures, and the number of protein folds that exist in nature must be relatively limited (269, 270). This technique has been widely utilized by many drug discovery which are based on the structural elements of the protein and has been proclaimed as the best prediction method of structural homologous protein (271).

3.2 Materials and Methods: Development and Application

Multiple *in silico* techniques are discussed here that are widely used to streamline and rationalize drug discovery efforts and aid in decisions of the drug discovery pipeline. Briefly they will be introduced before the specifics of the materials and methods are mentioned below.

Homology modeling involves the generation of a novel structure based on two primary elements: the sequence of a target protein and the model of an existing

structure with a similar sequence. It is widely used when x-ray or NMR structures are not available of certain protein but other family/species member proteins are available. Molecular dynamics (MD) simulations is an all-atom computer simulation of a system to study the physical movements of systems of atoms. Usually the atoms have their forces and potential energies modeled in a trajectory and the system is allowed to progress in small movements and the energies and potentials (force fields) are re-evaluated after each step. Both these programs are widely used in computational drug discovery efforts (268, 272).

This is a computationally expensive process, to simulations were carried out via the Texas Advanced Computing Cluster (TACC) in Austin. In order to determine what settings would yield the best performance on the TACC, a series of identical jobs were run to determine the optimal number of CPUs and nodes to utilize that would provide the longest simulation in a 24 hour time period (**Table 11.** Influence of Number of CPUs utilized on length of MD simulation).

Number of Nodes Utilized on TACC	Number of CPUs Utilized on TACC	Length of 24hr simulation (picoseconds)
1	144	36590
2	288	9649
10	1440	886
11.6	1680	831

Table 11. Influence of Number of CPUs utilized on length of MD simulation

Here we can see that the cross-talk between even just two nodes is a limiting factor, and therefore 144 CPUs is the optimal environment for this system to yield the best results before the simulation reaches the 24 hour wall time in the TACC environment.

3.2.1 Skp2 N-terminal Tail Homology Modeling

The initial coordinates of Skp2 was obtained from the Protein Data Bank (PDB code: 2AST) (239). The Skp2 N-terminal tail used PDB code 2OVP (273) as the template, which is based on FBXW7, a similar F-box protein. This structure has not the entire tail, but residues 65-95. Additionally, when these structures are aligned, the extra residues on 2OVP form a well-structured helix behind the H2 and H3 helixes of the F-box domain of Skp2. The alignment was performed using the UCSD Biology Workbench with the default parameters, and MODELLER (version 9.17) was used to build the homology model (274) and perform loop refinement. 10 models were generated and evaluated based on their molpdf and DOPE scores as well as refined and validated visually (no obvious structural clashes/error).

3.2.2 SCF Complex Homology Modeling

Similar to the N-terminal tail modeling, the initial coordinates of the SCF complex was obtained from the Protein Data Bank (PDB codes: 2AST, 1LDK)(242, 245). The gaps that are missing are as follows: Cul1 is missing the N-terminal residues 1-16, a large loop region with residues 55-84, and a second loop region

between residues 148-157. For Skp1, a loop is missing between residues 70-91. Finally, and again similar to the above modeling, the Skp2 N-terminal tail used PDB code 2OVP (273) as the template, which is based on FBXW7. This structure has not the entire tail, but residues 65-95. The residues were added based on the *H. sapiens* Skp1, Skp2, and Cul1 sequence (Uniprot IDs: P63208, Q13309, and Q13616, respectively. MODELLER (v. 9.17) was used to generate templates, and the best models were selected based on molpdf DOPE scores. 10 models were generated for each "run" and 10 runs were created, resulting in 100 overall models. From each run, a different random seed was selected to ensure that each run was independent of the previous ones. They were then inspected for obvious clashes/errors/improper angles. The top 2-3 candidates from each run were selected that had both the best molpdf DOPE scores and did not have any improper chains being inserted inappropriately into other elements of the SCF complex (e.g. MODELLER occasionally inserted the N-terminal loop region of Cul1 in between the H2 and H3 helixes of Skp2). These top 2-3 models from 10 runs were then compared and the top model was selected.

3.2.3 Skp2 Mutants Molecular Dynamics

The 2AST structure was used and stripped of all non-Skp2 atoms (Skp1, Cks1, p27, waters, and benzamidine groups. Three structures were generated: Wild type (no changes), a Trp97Ala mutant, and an Asp98Ala mutant. PyMOL was used to perform the mutations, and MOE performed a small minimization step on the side chain to

relax the mutants. AMBER software package was used as the software for the MD and the analysis (275).

3.2.4 Skp2 N-terminal Tail and the SCF Complex Molecular Dynamics

All molecular dynamics simulations were performed by the AMBER14 software package(275). For Skp1-Skp2, Skp2, Skp2-ZL-25 simulations, the 2AST structure was again selected. For the complete SCF complex with loop and tail regions, the model from the work done in the above section was used. The ff14SB force field (276) was induced to outline the atomic interactions between the protein elements. Structures were inserted in the centroid of a cubic water box using TIP3P (277). The system was energy minimized using steepest descent algorithm. After minimization step, the system was heated with water at 300K, and a density step was performed. After the density step, the production simulation was submitted. MMPBSA was used to analyze the energy of the ligand binding. RMSD and other AMBER analysis and processing of results was performed by the PTRAJ and CPPTRAJ packages (278). All simulations were performed remotely on the Lonestar HPC cluster at the Texas Advanced Computing Center (TACC) in the University of Texas at Austin.

3.2.5 Cks1/Skp2/p27 interface Molecular Dynamics

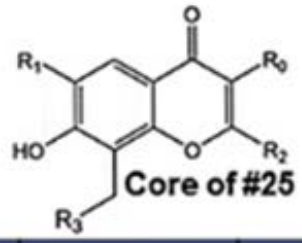
The 2AST structure was used, but only Skp1, p27 and the benzamidine groups were removed, keeping Cks1 and Skp2. All other settings used were same as previous MD runs described above.

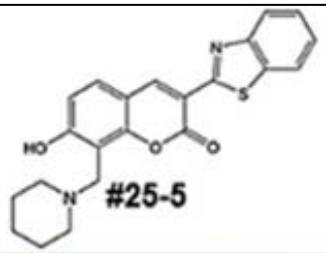
3.3 Structure Activity Relationship (SAR) Results

Next, a search was conducted for both lead ligands to find compounds with the same core scaffold, but had small modifications to uncover changes that would lead to better activity, namely, a structure activity relationship (SAR). A similarity search was carried out on each lead structure in the PubChem chemical database (279) (starting with 95% similarity) and structures that were commercially available were selected.

3.3.1 SAR results of ZL-25

For ZL-25, 14 structures were found and tested first as an in vitro Skp1-Skp2 binding assay similar to the primary assay used to discover ZL-25's activity (**Figure 35**).





ID	R_0	R_1	R_2	R_3	Activity
25	Benzothiazole	C_2H_5	H	$-N(CH_2)_5$ (piperidine)	Active
25-1	Benzothiazole	H	H	$-N(CH_2)_5$	Inactive
25-2	Benzothiazole	H	CH_3	$-N(CH_2)_6$ (azepane)	Inactive
25-3	Benzothiazole	H	CH_3	$-N(CH_2)_4$ (pyrrolidine)	Inactive
25-4	Benzothiazole	H	CH_3	4-N-Me-piperazine	Inactive
25-5	See #25-5 Structure Above				Active
25-6	Benzothiazole	H	CH_3	2-Me-Piperidine	Inactive
25-7	Benzothiazole	C_2H_5	H	4-Olamine-Piperazine	Inactive
25-8	Benzothiazole	C_2H_5	H	$-N(CH_2CH_3)_2$	Inactive
25-9	Benzoimidazole	C_2H_5	H	$-N(CH_2)_5$	Active
25-10	Benzothiazole	C_2H_5	-COO CH_2CH3	$-N(CH_2)_5$	Inactive
25-11	Benzothiazole	C_3H_7	CH_3	4-N-Me-Piperazine	Inactive
25-12	Benzothiazole	C_2H_5	CH_3	$-N(CH_2)_5$	Inactive
25-13	4-Me-thiazole	C_2H_5	H	$-N(CH_2)_4$	Inactive
25-14	4-Me-thiazole	C_2H_5	H	3-piperidin-3-ylpyridine	Inactive

Figure 35. Structure-activity relationship (SAR) of ZL-25 and its derivatives. #25-5 is illustrated separately due to its unique core structure.

The testing of these 14 revealed only two compounds with similar activity to the lead, listed as 25-5 and 25-9 (**Figure 36**). These were also tested in an in vivo ubiquitination assay and a multi-dose cell survival assay, and these two also

performed just as well as the lead, but not significantly better than the lead compound (**Figure 37**).

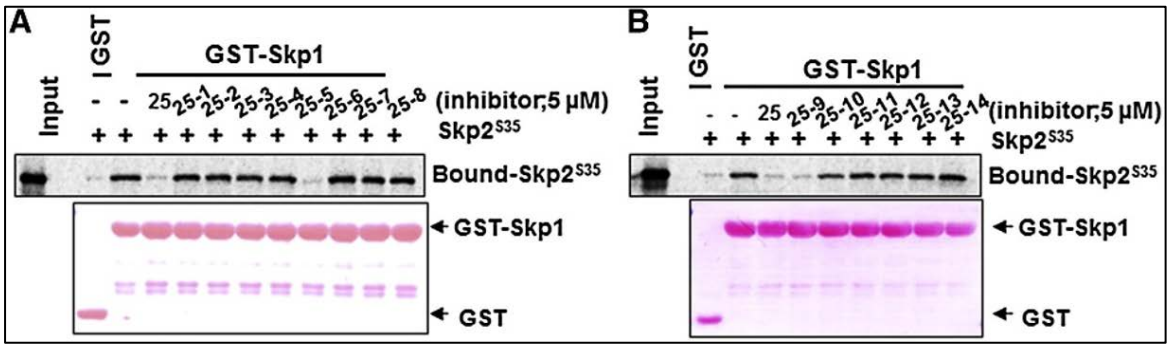


Figure 36. In vitro Skp2-Skp1 binding assay in the presence of DMSO, ZL-25, or its derivatives. Only 25-5 and 25-9 were as active as the lead, ZL-25.

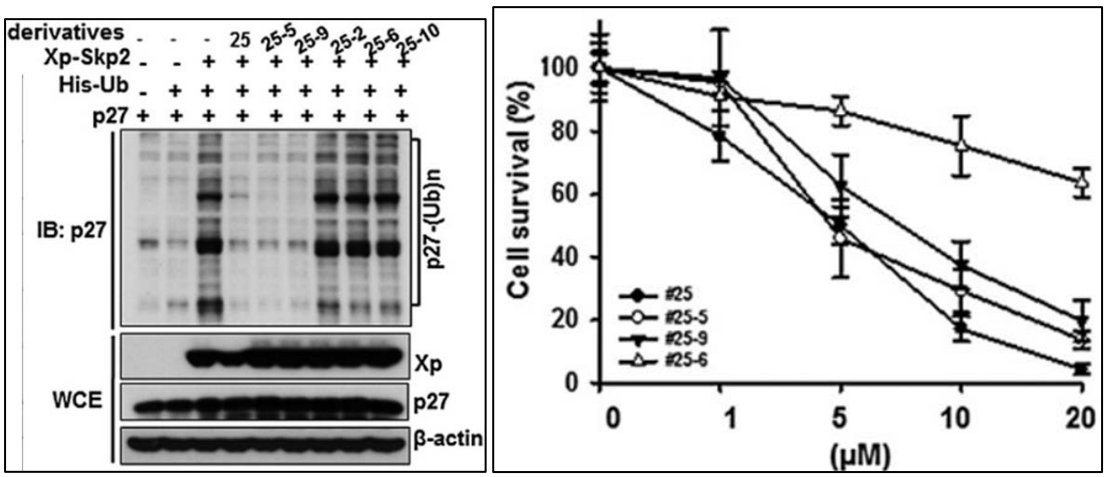


Figure 37. 25-5 and 25-9 results matched lead, but no better. 293T p27 ubiquitination assay with various constructs in the presence of DMSO or ZL-25 (Left). PC3 cells with various doses of ZL-25 or three other derivatives (25-6 is less effective) (Right).

Even though there were only two active compounds, the inactive compounds tell us about the structural features that are required and not required. Looking at Figure 35 at the chromone core structure and consider the different R groups, at the R_0 position the benzothiazole group (upper right side) seems needed; the 25-13 and 25-14 4-Me-thiazole lost their activity. The double ringed benzomidazole (25-9) and

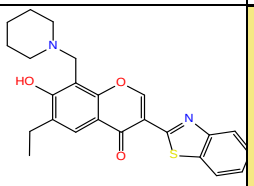
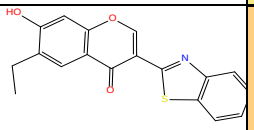
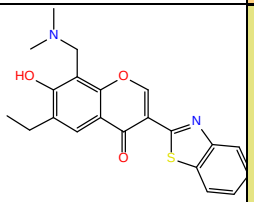
benzothiazole (all other compounds) groups are very similar and are able to more readily form pi-stacking interactions with Trp97 on Skp2. This also matches what we know about the mutagenesis efforts on the Trp97 group.

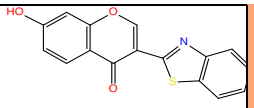
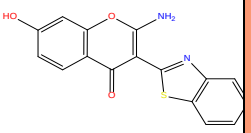
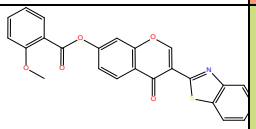
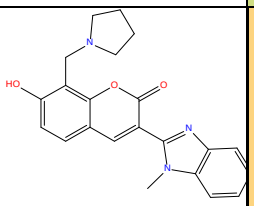
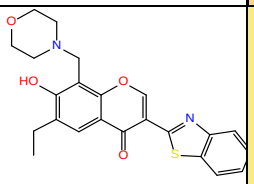
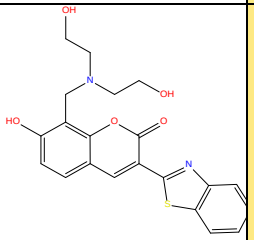
On the R_1 group (upper left), the ethyl group possibly is required, but at the same time 25-5, (that has a chromen-2-one core), was able to bind without an ethyl group, but this might be due to the hydroxyl group's interaction with the benzothiazole group in forcing it to not lie perpendicular to the chromen-2-one group. This also can be considered viable when one compares the structures with methyl or other large groups in the R_2 position (25-12, 25-10, 25-6, 25-2, and 25-3). The R_3 position appears to benefit from a large ring group.

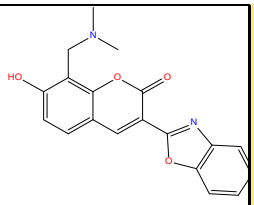
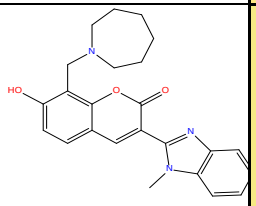
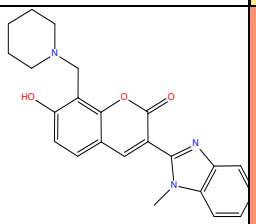
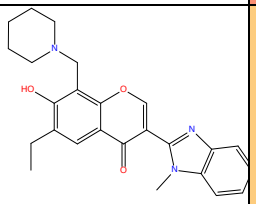
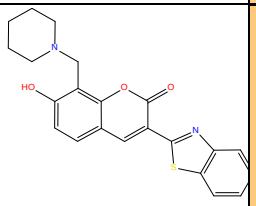
There have been additional rounds of SAR testing were carried out by three different sets of collaborators (the second round had 12 compounds, the third round had 11 compounds, the last one had six) and unfortunately the results often conflict with the previous group's work. The second round of testing used a similar pull-down assay to what was used in the first assay, but most of these results were inconsistent with the first round of testing. Additionally, compounds that appeared to perform well actually failed in secondary screens. However, the one conclusion that was definitive across both SAR tests is that a large R_3 group is required and cannot be hydrogen while there is no large group also on the 7' position (where the hydroxyl group lies on ZL-25).

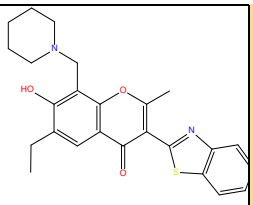
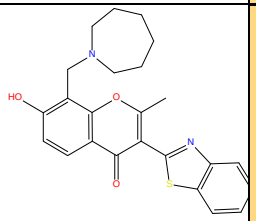
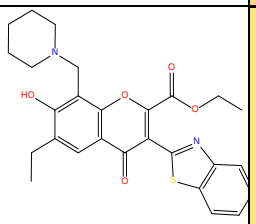
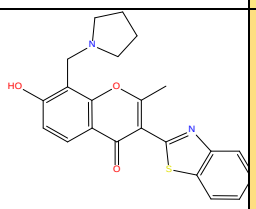
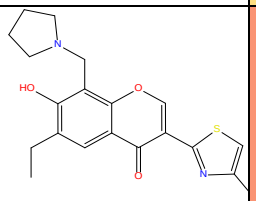
The results are shown below, all the tested compounds from the first two SAR studies were docked using three different scoring functions and the molecular weight

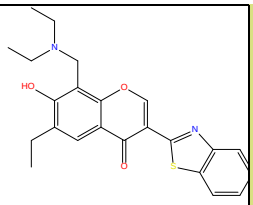
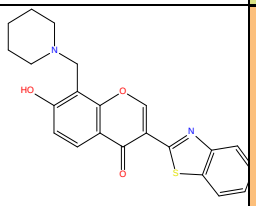
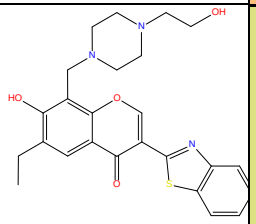
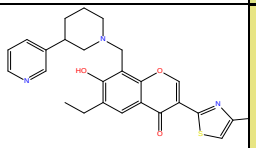
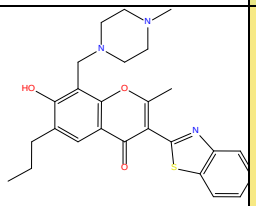
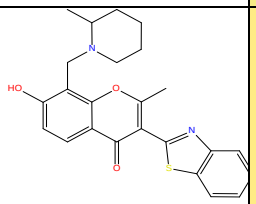
and logP displayed. The values are color coded relative to all the others. More favorable values are greener and less red. The column “Norm. w/ vehicle” indicates the activity in two ways: either a percentage value showing % of Skp2 binding to Skp1 (lower number is better), or a value of 5 through 1. This is a scoring system that was used to rank the ligands. Five is comparable to ZL-25, 4 is about 50% as good, 3 is even less as active as a 4 and was inconsistent in activity (experiments were performed in duplicate), 2 is even lower and inconsistent, and a value of 1 had absolutely no effect (**Table 12**).

ID#	Structure	Gold Score	ASP	Chem Score	Mol Wt	logP	Norm. w/ vehicle	Notes
5572358		56.71	11.68	9.906	420.5	4.58	10%	Original #25
5141378		53.89	14.42	13.68	323.4	3.6	100% (no effect)	No piperidin (R3)
5262653		57.85	13.51	14.49	380.5	3.61	8%	di-methyl R3

5363684		50.71	9.75	11.05	295.3	2.79	75%	No R1, R2
6942649		49.72	11.76	7.55	310.3	2.88	8%	amine R2, no R3
7354837		59.36	4.474	7.008	429.4	4.65	8%	replaced OH
6944343		54.32	17.53	16.16	375.4	3.95	35%	pyrrolidi R3
5575354		56.38	11.96	11.53	422.5	3.17	10%	Morpho R3
6942719		55.89	19.24	9.964	412.5	2.52	80%	di-ethyl amine R2

6944820		55.68	22.53	19.92	336.4	3.1	10%	di-methyl R3
6942580		56.69	18.13	16.04	403.5	4.84	8%	azepane R3
6241435		49.04	9.965	12.11	389.5	4.4	40%	positive control
5568376		53.66	10.41	7.553	417.2	4.09	Good_5	
6238284		53.47	13.85	17.16	392.1	4.89	Good_5	

5564702		54.8	9.776	7.612	434.2	5.31	Mild_4	
6239123		54.65	11.62	11.37	420.1	4.94	Mild_4	
6944596		55.6	8.116	9.606	492.2	5.29	Mild_4	
6240475		55.06	9.865	11.15	392.1	4.06	Mild_4	
9225029		49.37	16.75	14.57	370.1	2.75	Mild_4	

5359138		58.63	16.95	19.05	408.1	4.29	Cmpra bl_3	
5540697		52.71	17.68	16.4	392.1	3.77	Cmpra ble_3	
6942706		58.45	10.4	11.92	465.2	2.44	Half-Incnsist nt_2	
9242229		57.59	16.66	13.6	461.2	3.25	Half-Incnsist nt_2	
6944692		56.62	17.33	12.44	463.2	4.3	No Effect 1	
6238459		56.03	14.55	13.82	420.1	4.96	No Effect 1	

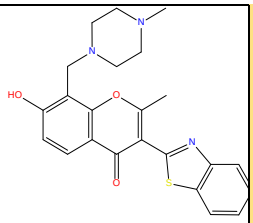
6239814		55.25	12.98	12.24	421.1	3.05	No Effect 1	

Table 12. Combined results of first two SAR studies. Values are color coded to have more favorable values in green. 5: comparable to ZL-25, 4 is about 50% as good, 3 is even less as active as a 4 and was inconsistent in activity (experiments were performed in duplicate), 2 is even lower and inconsistent, and a value of 1 had absolutely no effect.

The third round of SAR testing was a viability assay and actually used ZL-25 from two different vendors. The far right lane (#25) is ZL-25 from a primary vendor, #7 is from a secondary vendor and appeared to have slightly different activities. Two compounds were repeated across the two studies from the first round of SAR testing (data from **Figure 35**) and were consistently inactive: #6 is 25-5 from the previous study, as is #10 identical to 25-10. #12 is 25-13 from the previous study (**Figure 38**). Compounds from this study are in (**Figure 39**).

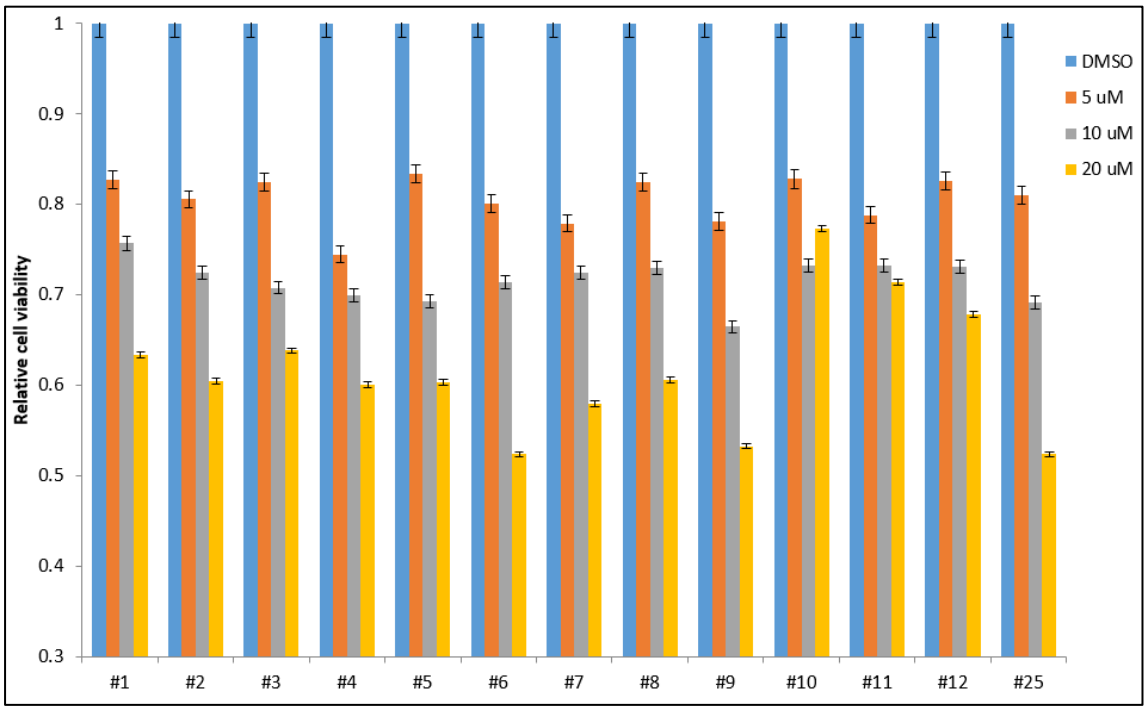


Figure 38. Second SAR study comparing viability. #7 is a batch of ZL-25 from a newer vendor, #25 is ZL-25 from a more previous vendor. Structures are visible inFigure 39.

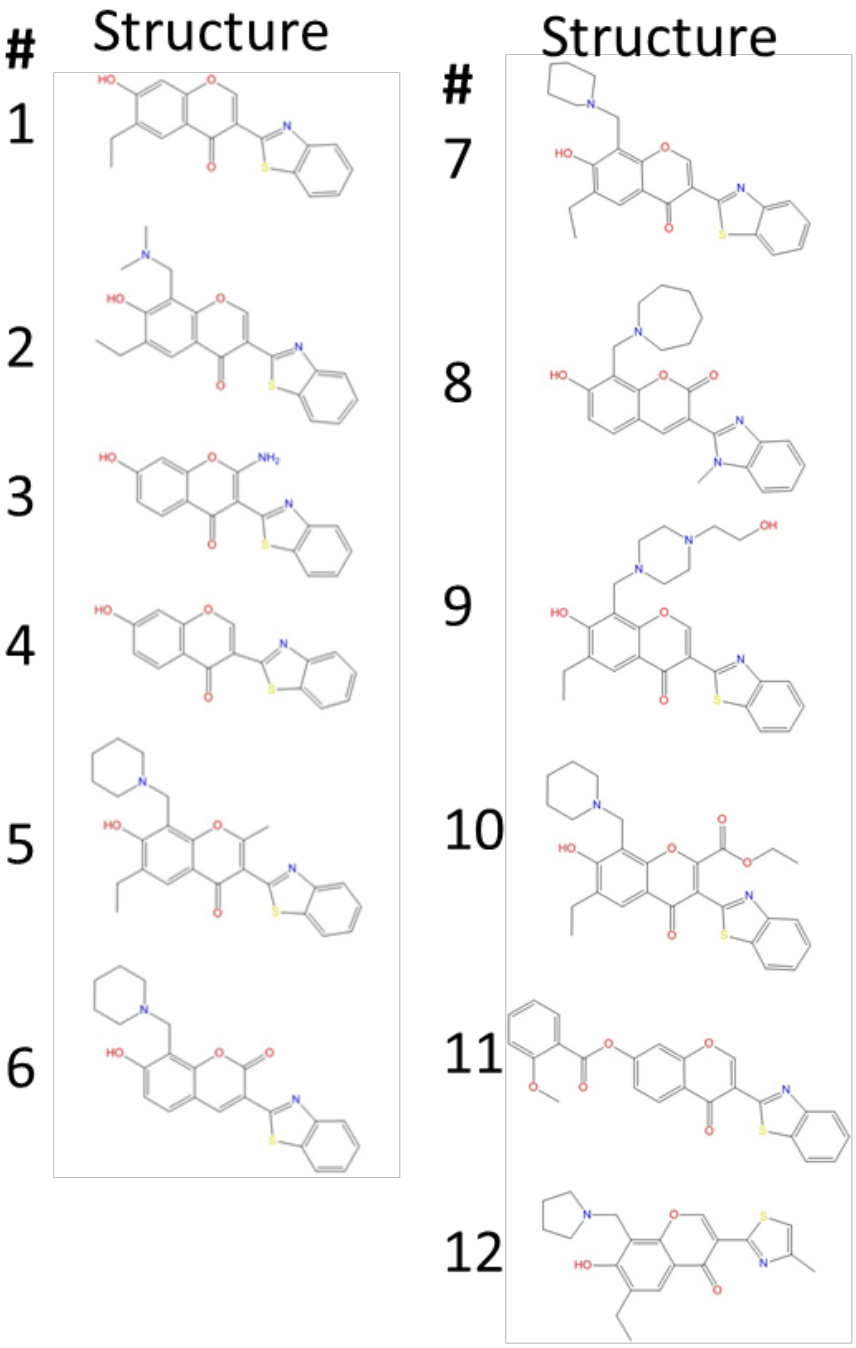
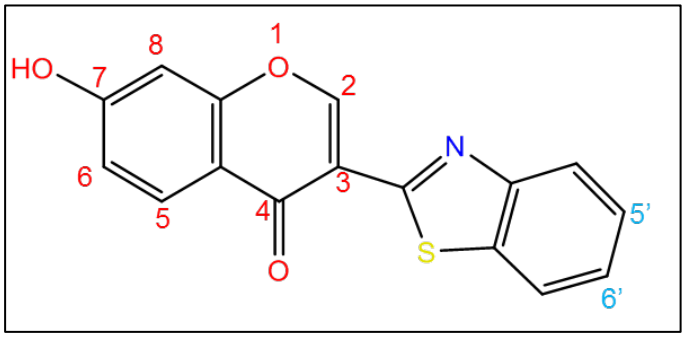


Figure 39. Compounds used in 2nd SAR study (results on Figure 38).

Compiling the results of all three group's work into one table has been performed, and the best understanding of the compilation of the work is presented here (**Figure 40**).

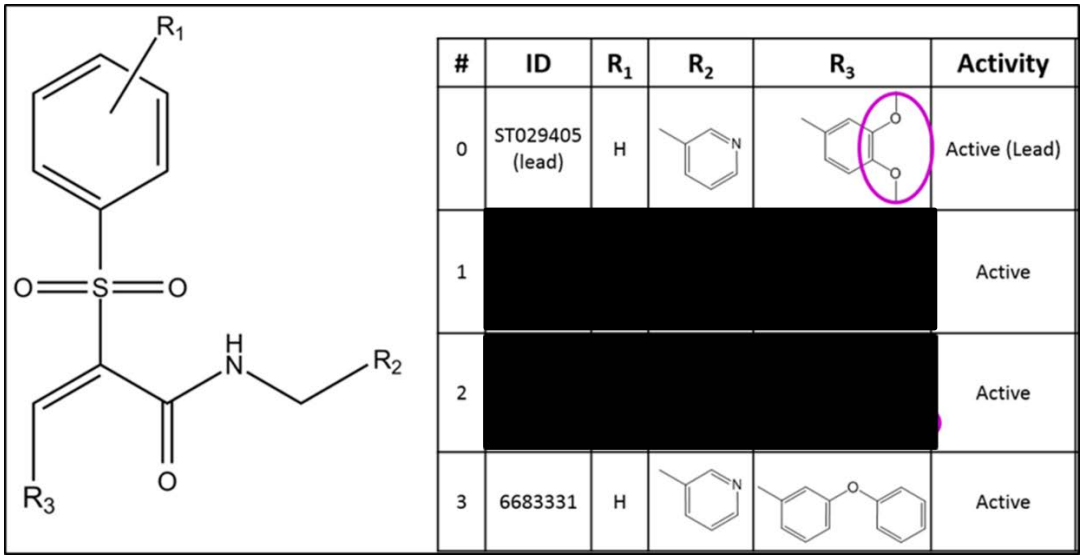


Position	Conclusions Drawn
2	–H is preferred , $–NH_{2\ is}$ also active, must be small group
8	Must be a relatively large group, or have a large group on 7
7 (-OH)	If no large group on 8, this can be a large, bulky group
6	Ethyl group not needed, but not clear what works
4 (=O)	Carbonyl group not needed here
S	Can be S or N-methyl
5', 6'	Prefer additional fused rings

Figure 40. Compilation consensus of conclusions of all SAR studies. On a per-position basis.

3.3.2 SAR results of ZM-405

ZM-405 has only had one primary SAR study, with only three compounds with one change on the benzene ring resulting in any activity (R_3 group) (**Figure 41**). The hits were tested in ECC1 cells for 24hrs with shown doses and after 24hrs, total protein levels were determined by western with specific antibodies and it was determined that 9276 and 8243 were considered better performers than ZM-405 (**Figure 42**). A new compound (which has its info redacted) became then the new lead compound candidate for this project and was renamed ZM-276.



#	ID	R_1	R_2	R_3	Activity
0	ST029405 (lead)	H			Active (Lead)
1	[redacted]				Active
2	[redacted]				Active
3	6683331	H			Active

Figure 41. SAR study of ZL-405 inhibitor. #2 (redacted) was shown to be more effective than ZL-405. Modifications to R_3 group highlighted.

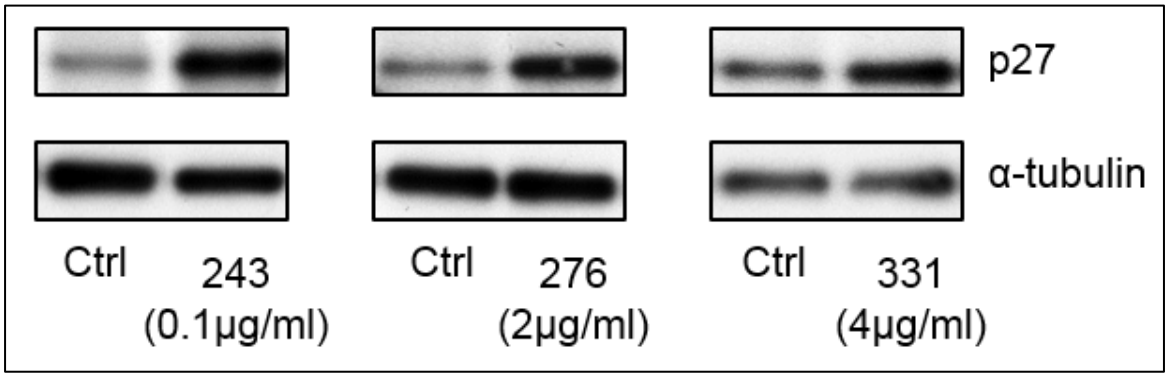


Figure 42. ECC1 cells treated with 3 ZM-405 analogs (last three digits indicate IDs). 243 (redacted) and 276 (redacted) performed better than ZM-405**.**

An additional 76 compounds were selected with high similarity to ZM-405 and were tested in a high-throughput assay for their ability to restore nuclear p27 levels. Unfortunately only the compounds with modifications on the R_3 group were active. There was great sensitivity on this group, and even other aromatic rings were completely inactive when compared to groups with large bulky modifications at the end of the R_3 ring (**Figure 43**).

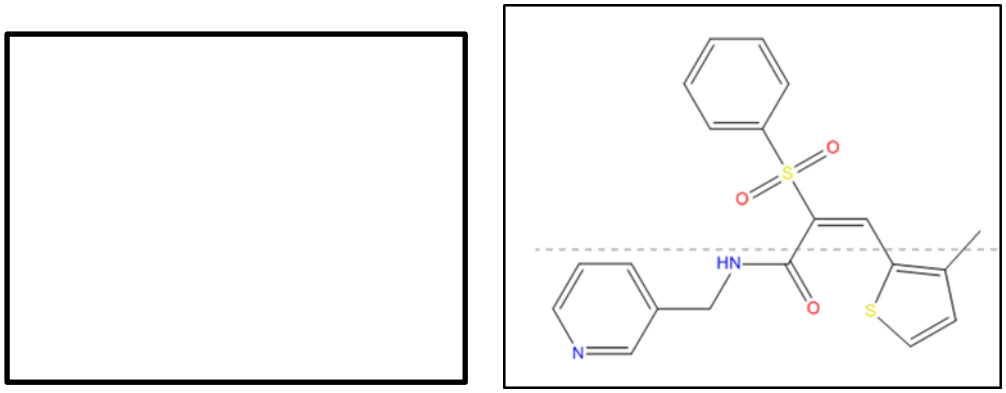


Figure 43. ZM-276 (left, redacted) was highly active in restoring nuclear p27, but SKT823113 (right) was almost completely inactive.

3.4 Homology Modeling of Skp2 N-terminus Tail Results

The original discovery of ZL-25 used the 2AST structure and the virtual screening took place on two regions of the Skp1/Skp2 interface. Originally, this project was to examine only the interaction between Skp2 and the ZL-25 ligand, perhaps taking into account Skp1 to explore the ligand binding and aid in the enhancement of the binding and efficacy of ZL-25. The ligand binding region of ZL-25 occurs on Trp97 and Asp98, and this site exists only two residues before the N-terminus end of the Skp2 2AST structure (Val95) and the rest of the N-terminal tail is unresolved.

As was said previously, the missing residues of the N-terminal tail have not been crystalized before the Val95 residue, and none have been described to have any known secondary structure using any method currently available. Using various prediction tools indicated that there might be some small secondary features. For example, PsiPred indicated that there is a small helix from residues 40-43 and a larger

one from residues 52-58, and a very small coil in 80-83, but the confidence in this coil is very low (**Figure 44**).

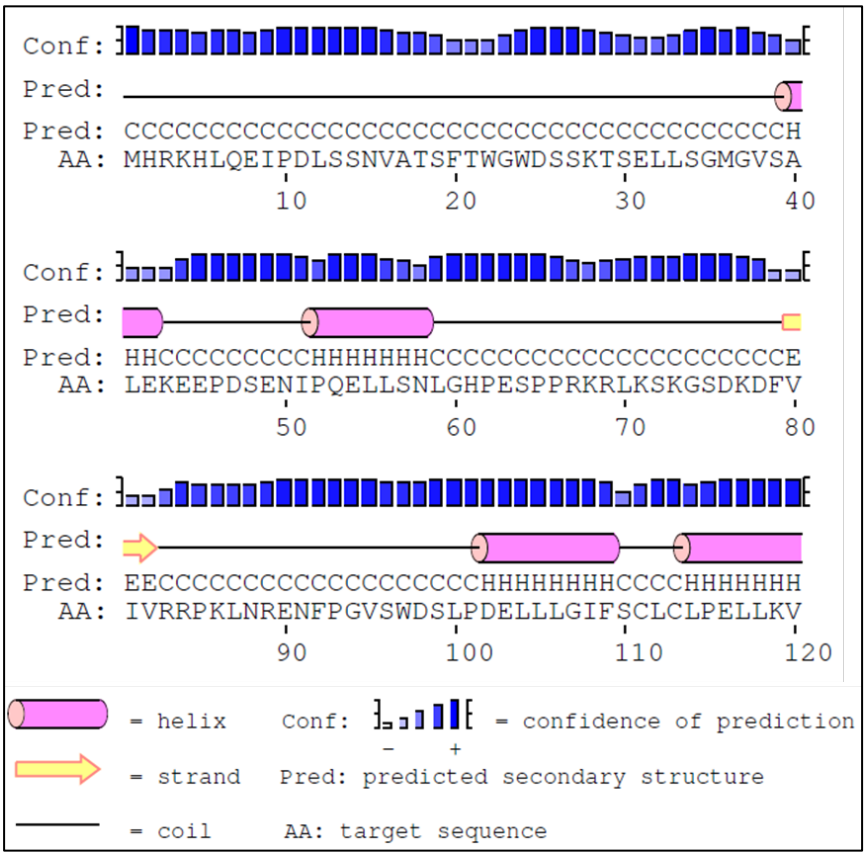


Figure 44. PsiPred prediction of secondary structure of Skp2 N-terminal tail. 2AST structure starts at residue 95.

The most adequate of an F-box protein template that was found to have the F-box domain and a longest region of the N-terminal tail was the 2OVP protein from Fbw7 (or FBXW7). This F-box protein had its additional 15-residue N-terminal tail existing both in a 2.5 turn (9 residue) helix and in a region closer to the LRR domain, and residing behind the H3 helix of Skp2, forming hydrophobic bonds between this tail and the Skp2 H3 helix (**Figure 45**). This 2OVP structure was initially used as the template on which to model a 15-residue extension of the N-terminal tail of Skp2.

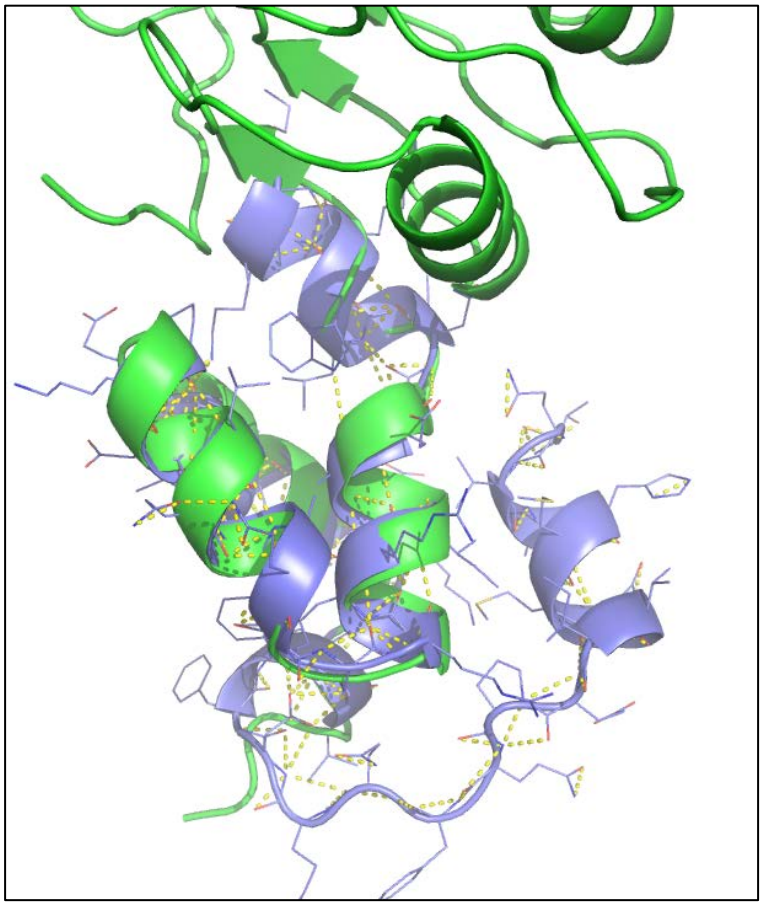


Figure 45. Modeling of Skp2 N-terminal tail. 2AST structure (green) compared to 2OVP structure (blue). 2OVP has additional residues crystalized and shown a position behind H3 helix.

Five final tail models were selected based on their molpdf and DOPE scores, and each were diverse in their final positions relative to Skp2 (this was not a selection criteria). They were called mod6, mod9, mod11, mod18, and mod20 (colored yellow, light blue, red, and purple, respectively in **Figure 46**). The highest scoring model was in fact mod11 (Brown structure in **Figure 46**), which adopted a configuration in which the N-terminal tail occupied the ZL-25 binding site and forms extensive contacts in between the H1 and H2 helixes of the Skp2 F-box which is the primary binding site of both Skp1 and the Skp1-Skp2 inhibitor ZL-25). This mod11 conformation might indicate that when a Skp2 N-terminal truncated mutant is

compared to WT Skp2, the mutant forms in the SCF complex almost 50% faster (data not shown).

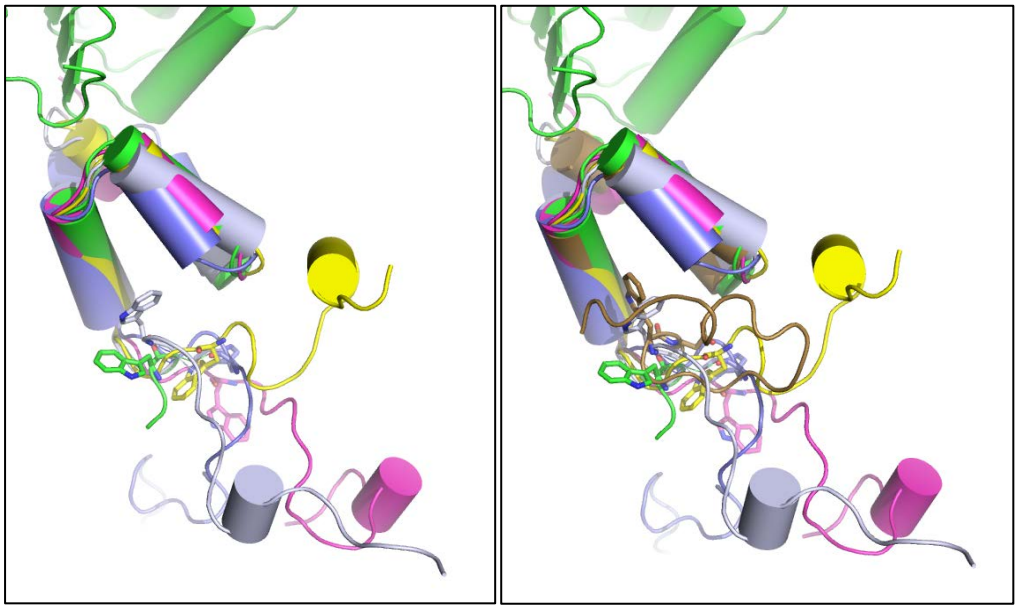


Figure 46. Highest scoring homology models of Skp2 tail. Four models showed on left have diverse structures (left). Mod11 (brown) is the only model that occupies Skp1 and ZL-25 binding site between H1 and H2 helixes (right).

3.5 Homology Modeling of SCF Complex Results

After further investigation into the entire SCF complex, it was determined that loop regions of Skp1 and Cul1 would have to be considered when modeling the N-terminal tail of Skp2, and other structures were added to the 2AST structure. The full model of the SCF complex revealed that the 2OVP structure is in fact inappropriate for the N-terminal tail model, as it lies in a region which is occupied by Cul1 and clashes with multiple parts of this large protein (**Figure 47**).

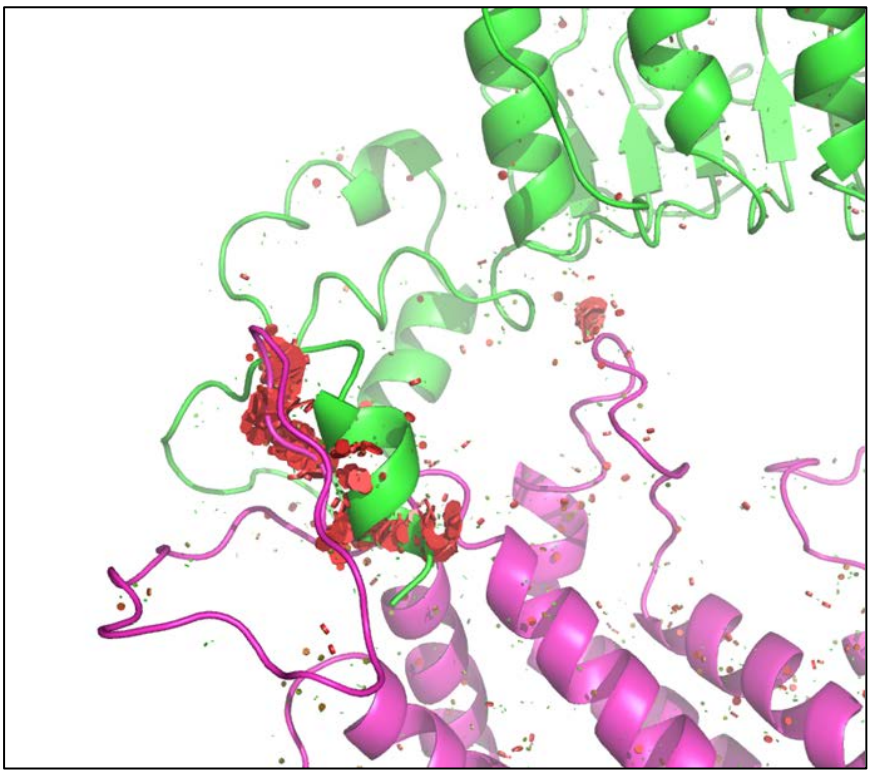


Figure 47. Skp2 (green) with Cul1 structure (purple) shows large regions of steric clashes (red). Using the 2OVP template of the F-box protein, this model cannot adopt a conformation into the SCF complex without clashing with Cul1 residues.

3.5.1 Active Recruitment Modeling of SCF Elements.

It has been understood that there are post-translational events that occur in on the SCF complex that facilitate the recruitment of its constituents to the SCF complex. In order to help understand how these events impact Skp2 recruitment, they were recreated in the current SCF model. It was recently discovered that the Ser256 residues in the LRR domain of Skp2 is important for Skp1 and Cul1 binding. (Ironically, SCF^{Skp2}, after Akt phosphorylation, binds to FOXO1 and induces its ubiquitination at the same residue on FOXO1, Ser256 (280)). This Ser256 on Skp2 is located at the end of the 4^{th} leucine-rich repeat (LRR4) opposite the side of where Cks1 binds and is oriented in a way that its side chain is pointing towards Cul1(**Fig5**).

This would indicate that, if Skp2 was able to “bend” towards Cul1, it would potentially interact with corresponding residues on Cul1. Merging the 2AST and 1LDK structures, the Ser256 residue on Skp2 is closest to Asp216 and Leu225 on Cul1, with distances of 19.4Å and 26.4Å, respectively, requiring only a 31.7° bend in the Skp2 F-box domain in order to come within the distance where these residues could form extensive interactions. The 10 residue loop region that was generated between Asp216 and Leu225 (DDAFAKGPTL) by MODELLER extends the loop towards Skp2 and the Ser356 residue. The most notable residue of this loop region is Lys221 in the middle, which extends its side chain even closer to Skp2.

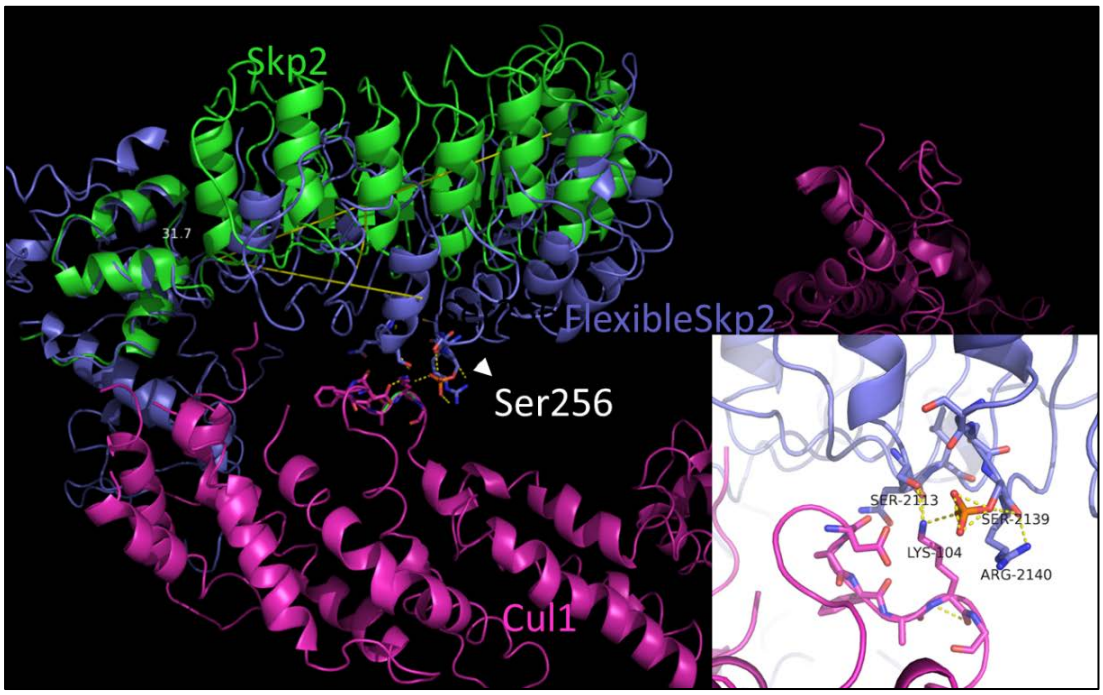


Figure 48. Active recruitment of SCF complex members is enhanced by phosphorylation of key residues. Ser256 on Skp2 results in its enhancement of formation of the SCF complex. Modeling of a key loop region on Cul1 shows that Lys221 on Cul1 can bind to Skp2 when the F-box domain bends only 31.7° (blue cartoon). The Cul1 Lys221A mutant displays similar rates of complex formation to a Skp2 Ser256A mutant.

When tested biologically, both residues are shown to impact the assembly of the SCF complex; mutation of Lys221 on Cul1 results in a SCF complex that is at

least 50% slower to assemble, and this mutation is similar to a Ser256 mutation on Skp2.

In terms of the post-translational modifications that result in active recruitment of the SCF complex, there have been multiple previous reports to this regard beyond those that take place on Skp2. For example, p27 is required to be ubiquitinated at the Thr187 and Ser10 residues (281). In Toxoplasma gondii, Skp1 has been shown to be glycosylated by five glycosyltransferases encoded by three genes on the Pro154 residue in an evolutionarily conserved pathway that results in a pentasaccharide at this Pro residue (282). This residue corresponds to Glu147 in humans. The assembly of the Skp1/F-box protein heterodimer is shown to be more efficient when this glycosylation is present, based on interactome studies (283). The SCF model presented here shows that this residue is not only solvent exposed, but in a structural position completely unprotected in relation to the other elements in the SCF complex (**Figure 49**). Additionally, all Cullin proteins (including Cul1) have been found to be neddylated by the ubiquitin-like protein Nedd8/Rub1 and gave the first clue as to how these large proteins are regulated(284). In regards to the Cul1 protein specifically, it has been shown to be neddylated by Nedd8 on K720 in its winged-helix B (WH-B) domain (although the authors report the residue as K696), which results in enhanced activity towards p27 (285, 286). To remove the Nedd8 protein, the signalosome protein COP9 signalosome (CSN) is required, and this allows the binding of paralogous regulatory factor proteins called Cand1 and Cand2(287). When Cand1 binds to Cul1, it results in a disruption of the F-box protein to bind to the Skp1 protein

and results in the inhibition of the conjugation of Nedd8(288). Additionally, Rbx1 (another critical component of functional SCF ubiquitination) is found to be necessary to require neddylation and also enhances the E3 ligase activity of the SCF complex. This is partially explained by the fact that the location of K720 on Cul1 is proximal to Rbx1 in the 1U6G structure (only 5.6Å away).

Figure 49. Post-translational events on other elements in SCF complex. A. Glu147 on Skp1 is glycosylated to enhance the binding of Skp1 to the SCF complex.

The complete model of the SCF complex with the Skp2 tail and all loop regions reveals the complexity of the system and potential interplay among the elements and their resulting biological aspects (**Figure 50**). The first 16 residues of the N-terminal region of Cul1 have the potential to interact both with Skp1, and with the ZL-25 binding site on the interface between Skp1 and Skp2. The large 30 residue loop (Ser55 through Val84) is even closer to Skp1/Skp2 (including the LRR regions of Skp2) and has an even higher potential of interacting with the ZL-25 binding site.

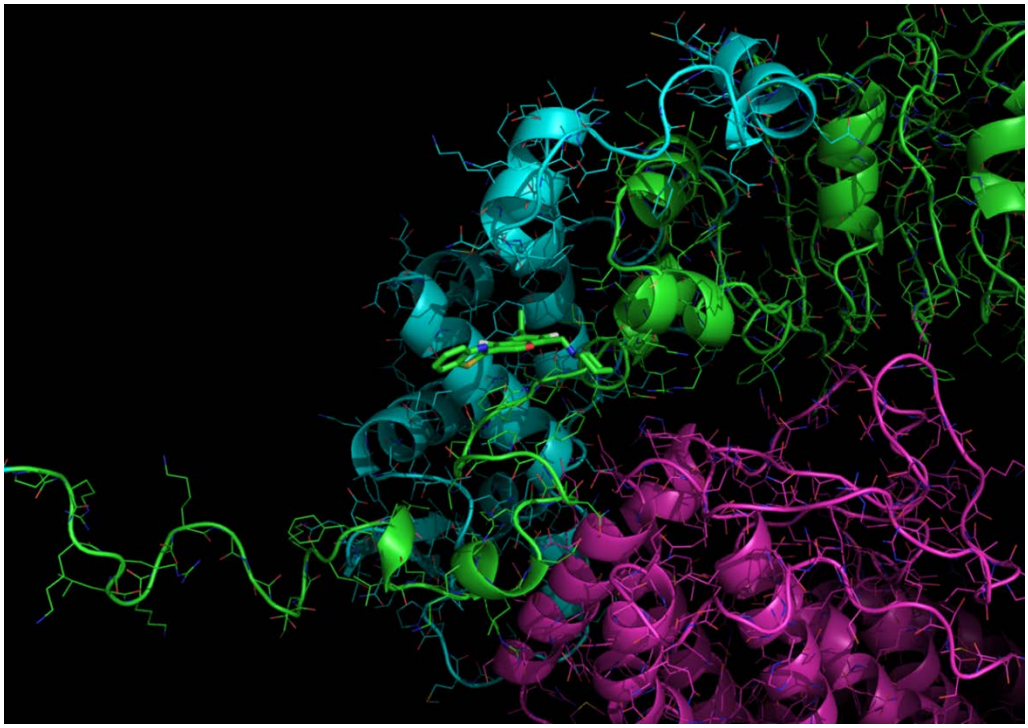


Figure 50. Generation of full model of SCF complex. Using Modeller9.17, a homology model was generated that includes missing residues of Skp2 N-terminal region (green), Skp1 including missing loops (blue), and missing N-terminus/loops of Cul1 (magenta). Additionally, the ZL-25 ligand was added in its docked pose (green sticks).

3.6 Molecular Dynamics of Skp2 Mutants Results

Simulation of the Skp2 protein over a long period of time (182ns) shows the relative stability of its different regions (**Figure 51**). The F-box domain is highly destabilized compared to the highly stable LRR region, and, surprisingly, the seatbelt region that extends from the end of the 10th LRR and loops back to the F-box domain shown a higher stability difference than the F-box domain.

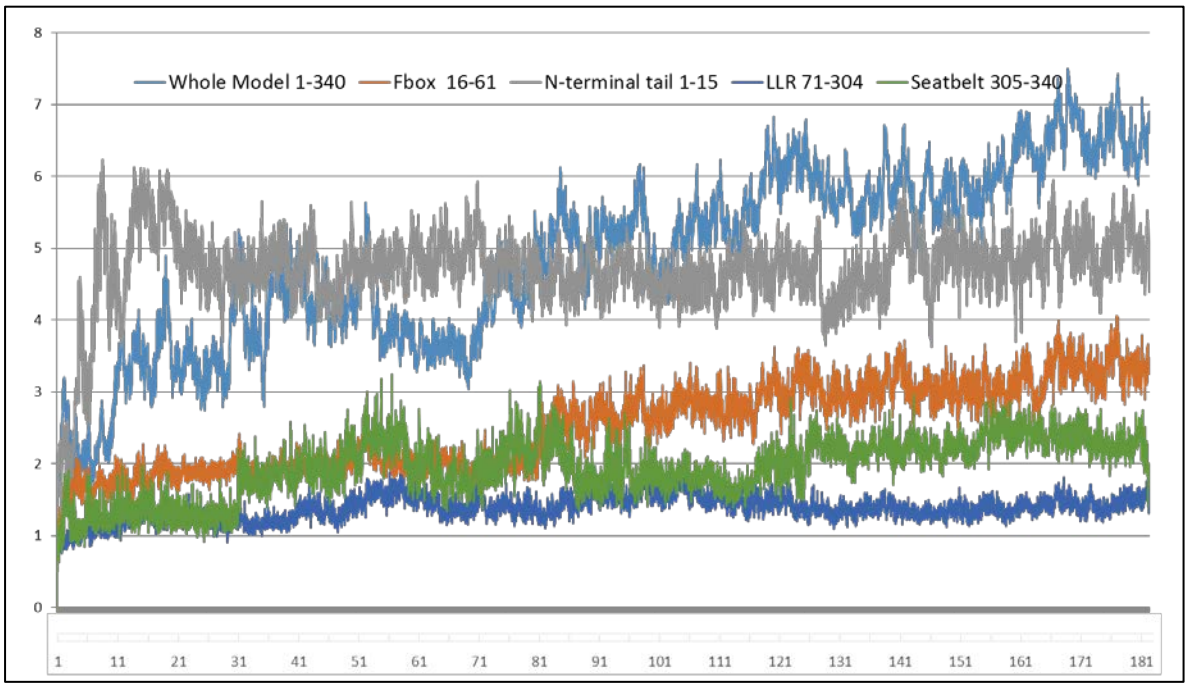


Figure 51. RMSD of backbone atoms from molecular dynamics simulations across 182ns of Skp2 with a 15 residue N-terminal tail of various portions of Skp2 represented. N-terminal tail and F-box show greatest amounts of variability, whereas the LRR and seatbelt regions are the most stable. The region in between the F-box and the LRR domain is very stable but has the potential to become destabilized for a short period of time.

Simulations of WT Skp2 in comparison with R126A and W127A mutants support the known biological information of these isoforms. It has been shown that the R126A mutant loses its ability to bind to Skp1 in the F-box domain, whereas the W12A mutant is able to retain its binding to Skp1 (176). When RMSD plots are generated for these three isoforms of the F-box domain, the WT and W127A isoforms show similar results, where the R126A mutant shows a significantly higher F-box backbone destabilization relative to the other two (**Figure 52**).

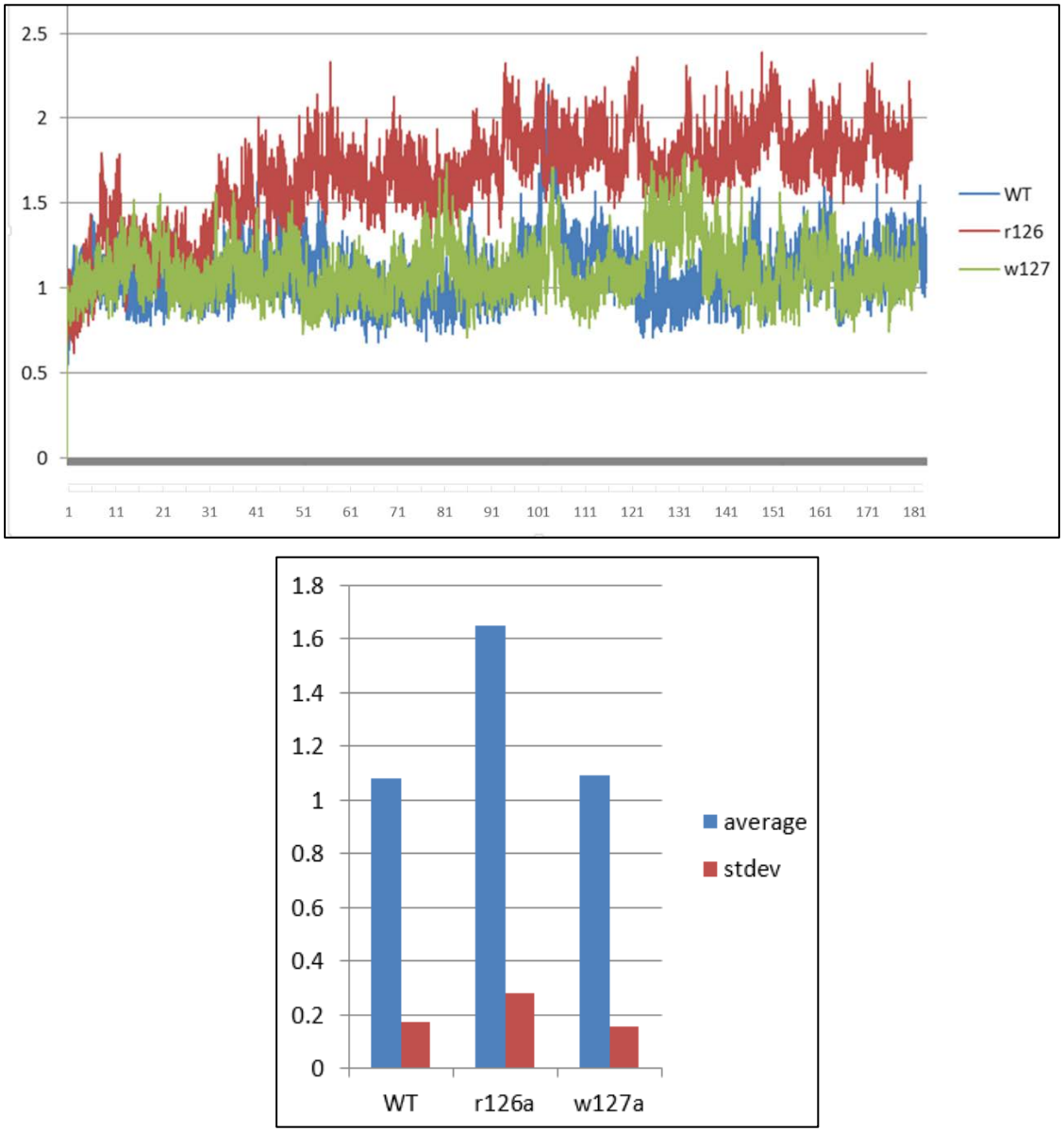


Figure 52. RMSD backbone plots of MD simulations of F-box domains of 3 different Skp2 proteins. (Top graph): 3 different simulations were carried out of Skp2: Wild-type (blue), W127A mutant (green), and R126A mutant (red). (Bottom): Average and standard deviation bar chart plot of the same data.

Upon closer examination of the F-box domains of these mutants, the H3 helix is the most shifted during this simulation, pulling apart the H1 and H2 helixes. Also, at times the seat belt region (the final residues of Skp2 that have folded back from the LRR domain), have a greater interaction with the end of the H3 helix.

3.7 Molecular Dynamics of Skp2 of SCF Complex Results

It has been widely understood that the flexibility of Skp2 occurs largely in its F-box domain, and in the linker region between the F-box and the first LRR region. This is flexibility is proven to be present upon examining MD simulations as well (**Figure 53**). This flexibility explains how it is reasonable to see how Skp2 and Cul1 can interact via a phosphorylated Ser256 residue on Skp2. In fact, only a 31.5° movement of the "hinge" region between the F-box domain and the first LRR region of Skp2 moves Ser256 to within 7.3Å of Lys221 on Cul1. This rotation is certainly possible, as two other structures of Skp2, 1FQV and 1FS2, have a rotation angle in the same area of 61.6°. A small rotation of the side chain on Lys221 brings it within hydrogen bonding distance of Ser256 on Skp2. This movement in Skp2 is seen primarily in three residues: Ser135, Leu136, and Trp137, which lie just after the H3 helix of the F-box and the first LRR region. This region is also able to move more freely as it is not significantly stabilized by the "seat-belt" region of Skp2 (residues 415-419).

Figure 53. 11 superimposed snapshots of Skp2 after 182ns simulations. The simulation shows very little movement in the LRR and 'seatbelt' region (upper portion), whereas the F-box domain shows significant levels of dynamics (lower portion).

There exist many post-translational modifications that are highly relevant for Skp2 that help us to understand the results presented here. The 95 residue N-terminal tail of Skp2 just before the F-box domain (also known as the nuclear localization signal, or NLS) is especially biologically highly significant for multiple reasons. The NLS lies within this N-terminal region, and Skp2 is acetylated at K68 and K71 by p300 (289). This acetylation stabilizes Skp2 via impairment of Cdh1-mediated proteolysis pathway, as well as promotes cytoplasmic retention. Multiple studies have shown that Akt binds to Skp2 in the N-terminal region (The substrate motif sequence of Akt is RXRXX(S/T), where X is any amino acid (290)), phosphorylating Skp2 on

Ser72 in a manner comparable to that of TSC2, a well-known Akt substrate. Other kinases (Akt2, SGK, or S6K) do not do this (214, 291, 292). It is understood that this event triggers SCF complex formation, E3 ligase activity, and triggered 14-3-3β-dependent Skp2 cytosolic relocalization (291). It is worth mentioning that subsequent studies found information to the contrary; Ser72 phosphorylation does not control Skp2 binding to Skp1 and Cul1, does not influence SCF ubiquitin ligase activity with Skp2, and has no affect Skp2's subcellular localization (293, 294). Therefore, when considering the effects of ZL-25 binding in the context of Skp2, one not only should consider the other elements of the SCF complex, but: the post-translational events that readily occur, the events that occasionally occur, and the resulting biological and structural implications of the interactions of these post-translational events in the context of all of the SCF complex.

3.8 Molecular Dynamics of N-terminal tail Results

The Skp2 N-terminal tail has been shown to have a high level of flexibility, but the other elements of the SCF limit the available space of the tail. When Skp1 binds to Skp2, the tail loses about 20% of its available inhabitable region (if we assume that it has complete freedom compared to the Skp1/Skp2 complex vs Skp2 alone). We know that, when tested biologically, the removal of the Skp2 tail greatly enhances the assembly of the SCF complex, so it is safe to assume that the N-terminal Skp2 tail can be in conformations where it prevents at least one of the other SCF members from binding. This is evident when we examine 4 snapshots of 1 simulation

of mod20 (one of the best performing tail models from the homology modeling of the tail residues of Skp2) over the course of 80ns. As the model progresses, it occupies 4 very unique positions with roughly the same secondary structure throughout. It is worth noting that this is not the full length tail of Skp2; it is only an extra 20 residues and not the full 95 residues (**Figure 54**).

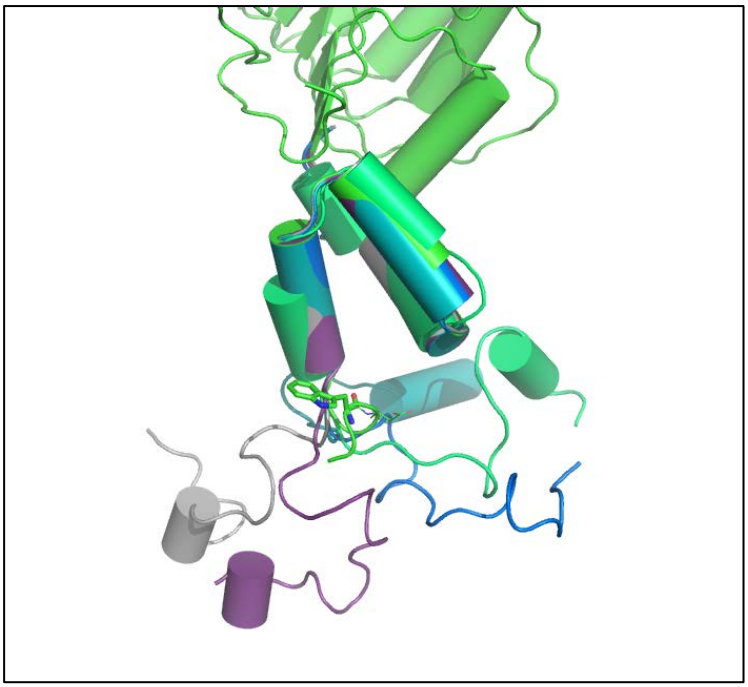


Figure 54. Four snapshots of MD of Skp2 model with N-terminal tail portion. 4 snapshots taken 20ns apart show very unique positions of the tail. Trp97 (green sticks) showing original 2AST structure.

However, the addition of the much larger Cul1 (making a Skp1/Skp2/Cul1 complex) restricts the position of the Skp2 N-terminal tail in a much more significant manner; it basically now has two primary positions that are 180° apart: one where the tail is extended away from the LRR domain of Skp2 (and away from the Cks1/p27 binding site) and is highly solvent exposed, or one where it extends in the complete

opposite direction towards the LRR domain of Skp2 (these two configurations will be referred to henceforth as simply exposed and buried).

The evidence that the exposed conformation is the most likely is to be found in the high number of post-translational events that occur on the N-terminal tail that require it to be solvent exposed. Akt1 (but not other kinases such as Akt2, SGK or S6k (214)) phosphorylate Skp2 on Ser72 (214, 291). Another study showed that the Ser72 phosphorylation event creates a priming site on Skp2 for subsequent phosphorylation by Casein Kinase I (CKI) on Ser75 which was detected by a cancer cell phosphoproteome (295). When both Ser72 and Ser75 are phosphorylated, association with Cdh1 is impaired, thus permitting Skp2 to avoid APC/Cdh1-mediated ubiquitination and destruction(296). This is significant, as Skp2 itself is degraded via the APC/Cdh1 ubiquitin ligase, and binds to the N-terminal domain of Skp2 on residues 46-94.One notable neighboring phosphorylation site to Ser72 also with notable effects is Ser64 (214). Interestingly, this site is the most conserved site phylogenetically within the region of Cdh1 binding and is found in all Skp2 orthologues, from vertebrates to insects, and also notable is the fact that this phosphorylation site is always located about 30 residues N-terminal to the F-box (292). One study found that mTORC1 also functions to regulate Skp2 by Ser64 phosphorylation and this actually has the potential to represent an oncogenic event in gastric cancer, as the combination of p-Skp2 and p-mTOR (phosphorylation of mTOR) expression was a better predictor of survival than either factor alone (297).

Another recent discovery of the N-terminal tail of Skp2 is its inclusion of a "destruction box" (D-box) motif. The core sequence of this motif is an arginine followed by two residues, then ending in a leucine (RXXLXXXXN/D/E) (298), which starts on Arg84 on Skp2 and ends on residue 92. Proteins with a D-box have been shown to interact with another ubiquitin-protein ligase called the anaphase-promoting complex or cyclosome (APC/C), and the first 90 residues of Skp2 are required for Skp2 degradation by APC/C in the G1 phase (299). This D-box domain on Skp2 is very close to the start of the F-box domain (residue 94). Considering all of these post-translational events and other protein elements that bind to the N-terminal tail of Skp2, it is highly unlikely that all these modifications and events occur when Skp2 is not bound to other SCF complex members or buried inside its other SCF partners. These facts indicate that the exposed form of the Skp2 N-terminal tail is the most likely conformation to exist biologically.

3.9 Molecular Dynamics of ZL-25 Binding Results

Simulations of the Skp1/Skp2 complex with the loop regions added to both proteins and with ZL-25 docked were performed. The regions that showed the greatest displacement were the two final helixes on the C-terminus of Skp1, and surprisingly the F-box domain of Skp2 showed only minor movements (**Figure 55**). Also interesting, was that ZL-25 actually flipped the relative locations of the benzothiazole group and the chromone core compared to the original starting docked structure at the relative start of the simulation, at about 16ns into a 32ns simulation.

This rotation was seen on at least 2 separate simulations (**Figure 56**). This is actually still reasonable, as the chromone core still has aromaticity and can still form pi-stacking with Trp97 (albeit not as strong as the benzothiazole group), and the Asp98 residue can potentially interact with the benzothiazole group.

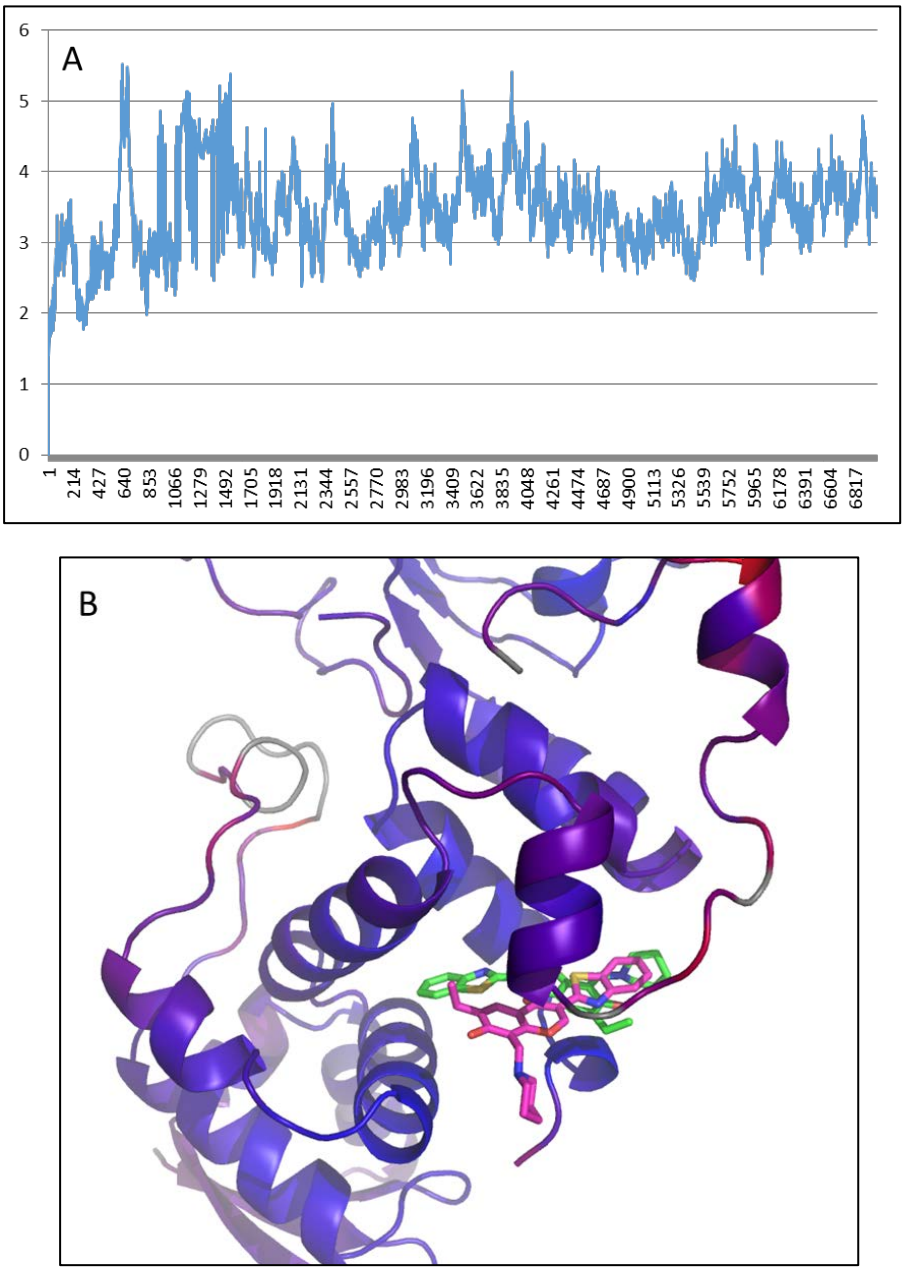


Figure 55. Results of MD simulation of ZL-25/Skp2 complex **A**. Backbone RMSD (blue) of Skp1/Skp2 in complex with ZL-25 of a 140ns simulation. **B**. RMSD color-coded comparison of starting structure with final structure. Color goes from blue (small/no displacements) to red/grey (large displacements). ZL-25 shows a rotation across an axis perpendicular to main axis of Skp1/Skp2 structure.

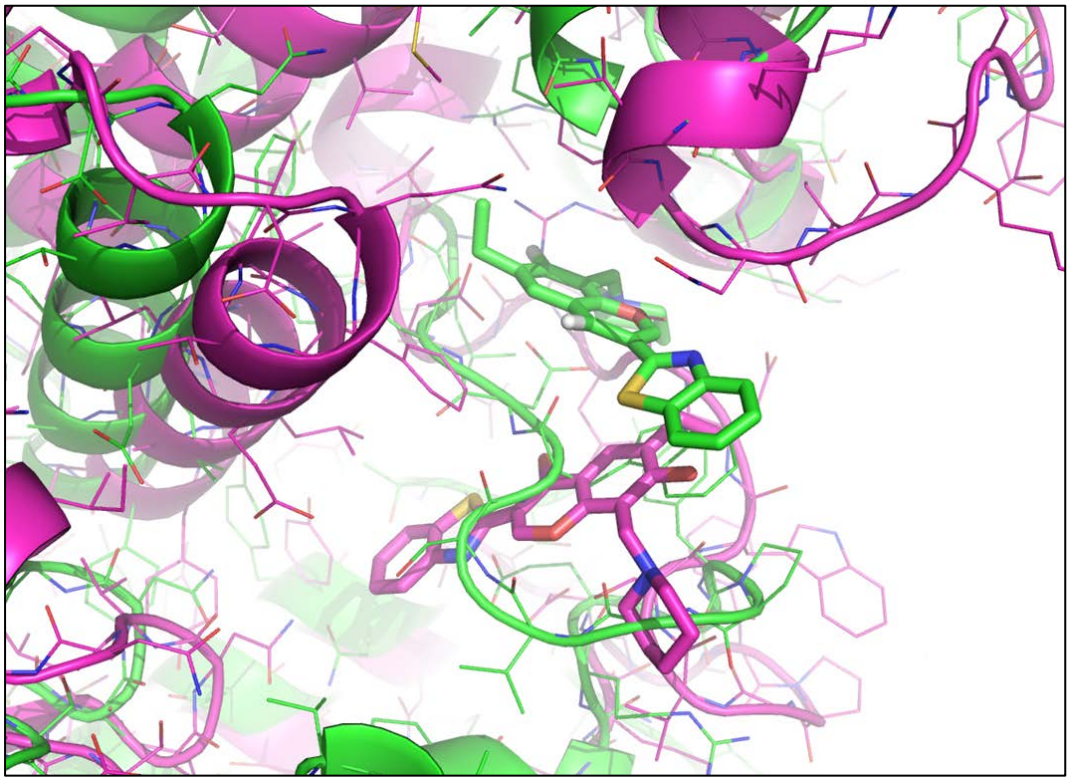


Figure 56. Molecular Dynamics simulations of ZL-25 ligand in complete SCF complex with Skp2 N-terminal tail. From the starting structure (green), ZL-25 rotates its orientation relative to the starting structure after approximately 16ns (magenta).

The average RMSD from the starting structure (Skp1/Skp2/ZL-25) to the end of the 140.3ns run was 2.95Å, but the maximum distance was 10.21Å. Additionally, comparing the per-residue RMSD of both the bound and unbound forms of Skp2 with ZL-25 show a destabilization in the F-box domain, whereas the other regions of Skp2 remain basically unchanged (**Figure 57**).

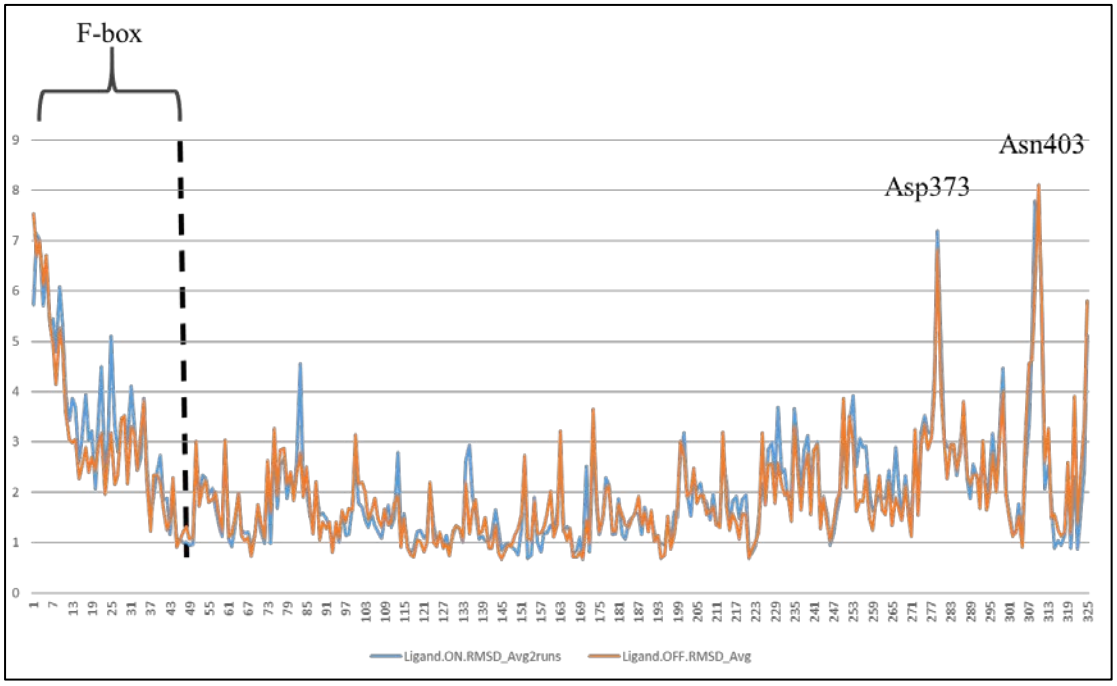


Figure 57. Average RMSD per-residue of Skp2 in both bound (blue) and un-bound (red) states. Residues Asp373 and Asn403 are in highly variable regions of Skp2 F-box shows highest destabilization effects upon ligand binding. Asp373 and Asn403 mark the end of the LLR domain, and the middle of the seat-belt region, respectively.

This F-box destabilization is also indicated in the Ramachandran plots of the bound and unbound forms after a 42ns MD simulation. The majority of the residues that show changes in their dihedral angles lie in the F-box domain and there are in fact more than three times as many in the bound form vs the unbound state (7 vs. 25). The two major improper angles (red residues) in the starting structure and the lone improper angle in the bound form are either residues on the edges of the LRR domain, the end of the LRR transitioning into the seatbelt, or in the middle of the seat belt; residues that are highly solvent exposed or in regions where high torsion will naturally exist (**Figure 58**).

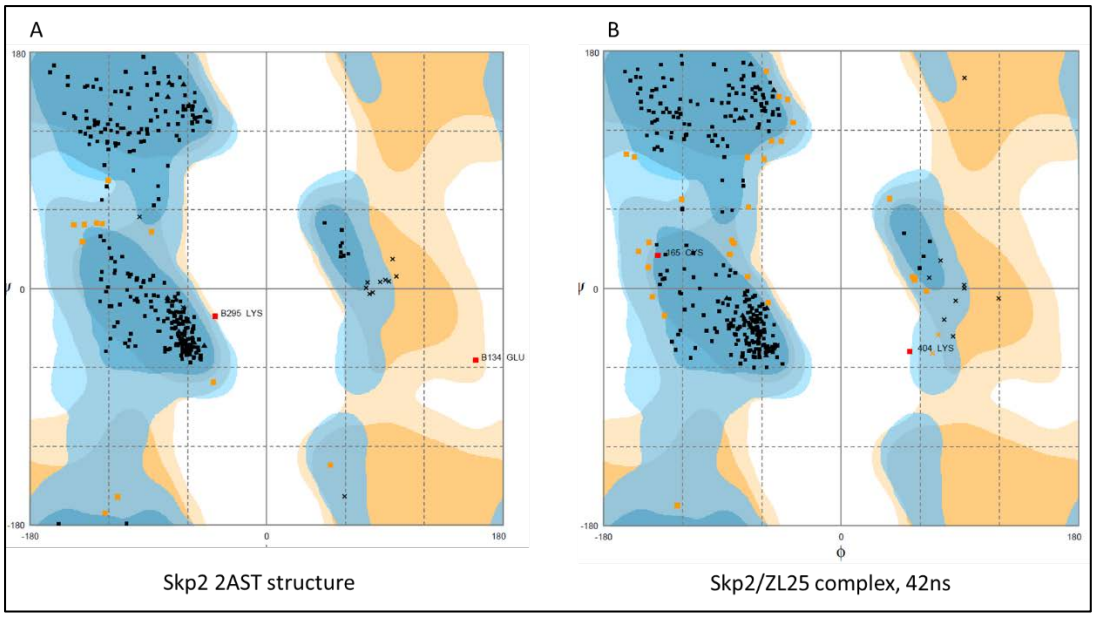


Figure 58. Ramachandran plot of (**A**) unbound and (**B**) bound forms of Skp2 with ZL-25 ligand after 42ns MD simulation. The majority of the shifted angles are from the F-box domain.

It has been estimated that 85% of proteins undergo conformational changes upon binding to a ligand, and these changes can range from a simple rotamers (300) of key binding site residues to larger domain rearrangements (301). As ZL-25 is a protein-protein interaction disruptor, the results from the MD simulations here show that this inhibitor can destabilize Skp2 in its F-box domain in a manner that results in its ability to prevent its binding to Skp1. As Skp2 is the rate-limiting component of the SCF complex, its inability to bind to Skp1 results in a loss of ubiquitination activity for its ligands such as p27.

To summarize, in this section, we used homology modeling and molecular dynamics simulations to indicate that the intricacies of the SCF complex are not confined to just one protein and there are also post-translational events that must be considered when selecting models and evaluating results from these studies.

Chapter 4: Summary and Future Directions

4.1 Summary of Hot Spots and their role in Drug Development

This study offers significant evidence that hot spot residue analysis offers the potential to expand the drug-target space by allowing researchers to examine PPIs as potential targets. PPIs have been proposed to be good candidates for drug targets, and notable cases include examples such as iNOS, LFA-1 and 14-3-3 pathways (302). Targeting PPIs have unique drug discovery challenges that will need to be overcome if they are to become a viable alternative to targeting enzymes or receptors like GPCRs. Standardized enzymatic turnover or kinetics can often not be used for some PPIs as a surrogate for the potency of hits, so design and execution of proper assays to measure responses that fit the proposed mechanism must be carried out. Also, accurate prediction as to how PPI disruptors/inhibitors alter the pathobiology and function of target proteins will be more difficult than enzyme/receptor predictions, as these traditional targets commonly have distinct residues/regions that are catalytic in nature and can be focused upon(303). Finally, proper library design must be utilized and as more PPI disruptors/inhibitors are uncovered, perhaps pharmacophores or moieties will be uncovered that are more likely to lead to their activity in certain features of PPIs.

Alanine scanning mutagenesis, while costly and time consuming, can be carried out using the prediction tools carried out here, but care must be given as to the potential biases or weaknesses of each prediction method. When possible, ensemble

methods as well as simple structural analysis should be used to supplement these techniques to minimize false positives and false negatives.

The theory and practice of using small molecule inhibitors to disrupt protein-protein interactions of protein complexes is a novel one, and one that will require further exploration of the dynamics of proteins involved. While still in its relative infancy, proposed procedures and strategies have been outlined for using hot spots towards developing small molecule inhibitors of PPIs (304). This exploration is aided by the use of computational tools such as homology modeling and molecular dynamics. Advances in *in silico* high-throughput screening are also needed to identify active compounds and will speed the process of identifying hits. Multiple groups in both industry and academia are making use of high-performance clusters (HPCs, also called grid or distributed computing) that use multiple networked computers that can be set to perform intensive calculations (most notable are MD simulations and in silico virtual screens).

Also of significance to the advancement of this field is the ever-increasing computing power and capacity of individual computers in both hardware and software. As computational speeds increase and molecular dynamics simulations increasingly become optimized in their scoring function accuracy and speed to take advantage of elements such as parallel computing and utilization of both next-generation graphical processing units (GPUs) and CPU architecture(305), these simulations can be performed on timescales that will allow for the visualization and analysis of studies such as large, multi-subunit domain movements, assembly of large

protein complexes, and protein-ligand interactions such as the ones presented here, even the potential of simulations heading into the second timescale (as opposed to nanosecond or millisecond) are predicted to be capable based on current trends by the year 2022 (306-308)

Hence, we believe that a well-developed structure-based modeling technique designed specifically for PPIs, e.g., structure-based virtual screening, is still in its infancy but holds the potential for increasing the drug-target space for cancer and can potentially be expanded towards other diseases.

4.2 Skp2 and the Ubiquitin pathway as a Potential Clinical Target

Skp2 is a prime target for further development as a cancer target, as it clearly shows over expression (and not mutation or truncation/deletion) in multiple cancer types. Compared to bortezomib, the general proteasome inhibitor, targeting the E3 ligase allows for tighter specificity and selectivity with less associated toxicity. There exist multiple endpoints for targeting the ubiquitin pathway, and this study gives strong evidence that there exist multiple mechanisms by which disruption of the SCF complex can be achieved: targeting the large interface of the Skp1/Skp2 interaction in the F-box domain of Skp2, as well as targeting the p27 binding site on the Cks1/Skp2 interface. As we have shown with the inhibitors presented here, this specificity and selectivity can be achieve; other highly similar F-box proteins were not active towards ZL-25 (276 has not been tested against similar proteins at this time). As has been discussed previously, only two compounds (Lenalidomide and Pamalidomide) that

target elements of the ubiquitin pathway have been FDA approved (excluding the proteasome inhibitors). When one considers that there are approximately 40 E2 (ubiquitin-conjugating enzymes) and over 600 E3 enzymes in humans (309), clearly there exist multiple opportunities to utilize PPI disruption (especially when considering hot spots) to rationally design small molecules to other elements of the ubiquitin pathway (Figure 59).

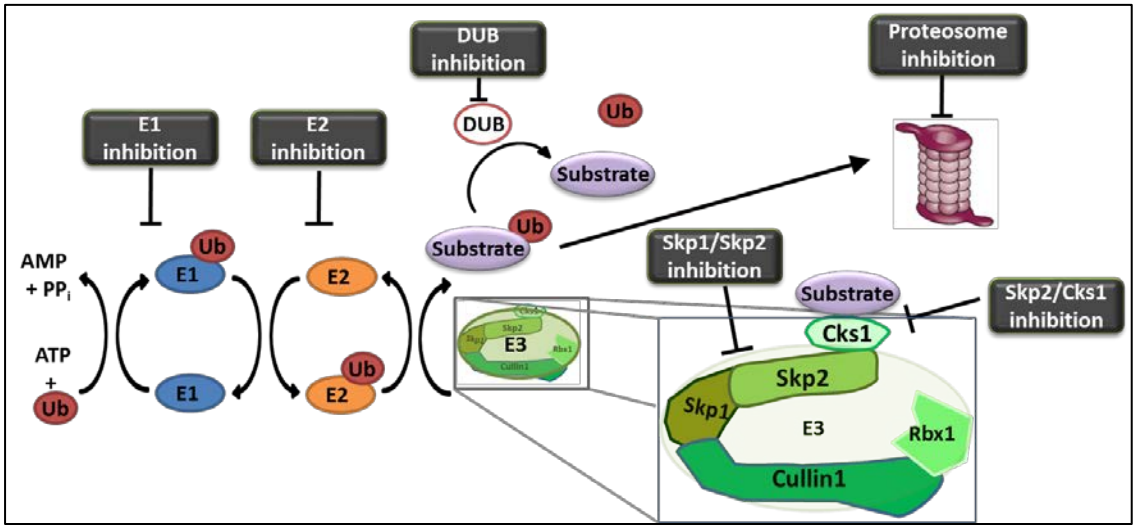


Figure 59. Potential strategies for targeting the ubiquitin/proteasome pathway. This study presents two strategies for targeting the E3 ligase system of SCF/Skp2 (green).

To look for other potential projects on which small molecules could be designed for small molecule inhibition similar to the two studies that were performed on the two regions of Skp2, a literature and structure search was carried out for all other E3 ligase proteins, there exist a few potential targets that should be considered (**Table 13**). Most notable is FBXL10 and FBXW7, both have evidence of pathology to cancer, and have structures available. Some candidates do not currently have crystal structures available; for these either homology models could be built, or as crystallographers improve, these structures might be resolved at a later date.

F-box protein	PDB codes	Fbox crystalized	E3 ligase	Implication in cancer	KO model cancer evidence	Pathological evidence in cancer	Major Substrates	PubMed entries	Good candidate for targeting?
FBXW1 (β-TRCP1)	1P22, 2P64	**Y**	y	Context Dependent: NFKB activation, WNT signaling?	y	BOTH overexpress. AND deletion	β-catenin, Mcl1,c-Myc, Cdc25A, Wee1A, CyclinD1, Mdm2	FBXW1:6 βTRCP1:66	Not really: not a clear oncogene, and isoforms are functionally redundant
FBXW7	2OVP, 2OVQ, 2OVR	**Y**	y	Tum.Supp: Degrades several proto-oncogenes. Mutants found in : ovarian, breast, CRC	y	mutation	CyclinE, c-Myc, c-Jun, Mcl1, Hif1α, mTOR, B-Raf, Enolase1	459	Widely studied previously
FBXL1 (SKP2)	**1FQV, 1FS1/2, 1LDK, 2ASS, 2AST**	**Y**	**Y**	**V. Strong evidence as oncogene**	**Y**	**overexpress.**	**p21/p27**	1003	Yes, being targeted now
FBXL3	4I6J	**Y**	y	Emerging TumSupp., KO mice show altered circadian rhythm, Mutants found in some cancers	**Y**	mutation	CRY (circadian clock)	27	**Not ideal: Tumor suppressor, its substrate is not significant**
FBXL5	**3U9J, 3U9M, 3V5X, 3V5Y, 3V5Z**	no, hemerythrin domain	y	Undetermined	??	??	p150:binding to dynein and microtubles IRP2:maintenance of iron concentrations	28	Not ideal: unknown implication in cancer
FBXL10	4O64	no, zinc fingers	part of PRC1 cmplx	**Emergine Oncogene; has dual functions as demethylase and Ub**	**yes**	**overexpress.**	p16(Ink4a/4b/4c), H2A, c-Fos	FBXL10:27 KDM2B:43	**Yes, but no Fbox and dual functions might be 2x edged sword**

Table 13. Potential F-box protein targets for future examination of small molecule inhibition.

The stabilization/restoration of p27 appears to be the driving force behind the effectiveness of both SCF complex inhibitors, and this has been demonstrated previously to be effective and achieved by other methods. One study has revealed two significant findings about dasatinib: first, that it enhances paclitaxel-associated apoptosis by increasing p27 expression, and second, that dasatinib induces nuclear p27 expression specifically by inhibiting its phosphorylation on Thr187 (which is necessary for it to bind to the Cks1/Skp2 complex and leads to its 26S proteasome degradation), as well as Ser10, and Thr157 (310). Each of these residue phosphorylation events have been reported to control the stability of p27 (reviewed in (311)). This second study also discusses the fact that anti-HER2 antibodies (such as trastuzumab) upregulate p27 protein levels in breast cancer cells that overexpress HER2 but decreasing Thr187 phosphorylation of p27 (312). This is in addition to the

at least six signaling targets and pathways that are modulated by trastuzumab. Another study proposed that the cisplatin analogue and DNA-damaging agent DAP is able to induce p27 via a novel mechanism. The authors propose that basal levels of p27 are recruited into two complexes: CDK4 and CDK2. The CDK2 complex results in phosphorylation that leads to rapid degradation of p27. They then propose that it is in fact p21 that is induced and is able to inhibit CDK2 and prevent p27 phosphorylation, thereby stabilization of p27 via preventing its phosphorylation in a p53-dependant manner via DAP induction (313).

These studies reveal the paradigms of p27 induction, and when they are compared to the current study of using Skp2 inhibitors in two different regions, the differences are worth noting. First, there is direct binding of the compounds to Skp2, (indicated by the mutagenesis experiment) and there are no overt toxicities seen in the mouse xenograft studies. Second, as the ligands phenocopies the effects of Skp2 genetic deficiency, cellular senescence is achieved in a p53-independent manner (176). This is significant as advanced cancers often develop resistance to treatments that induce p53-dependent apoptosis and senescence.

4.3 Future Directions in ZL-25 Development

Here we present novel inhibitors of two different regions targeting different components/regions of Skp2 and its role in the SCF complex. Both regions were analyzed using hot spot prediction techniques and the suggested hits were discovered via virtual screening and were later proven to have biological activity in line with the

described mechanism. For the Skp1/Skp2 interface, an initial ligand set of over 56,000 compounds was used, and from using two unique docking scoring functions, we were able to produce a list of 25 compounds that we predicted would have biological activity. Both sets of lead inhibitors are in the pre-clinical stage, and there is significant work and study that is needed in their development to advance them into clinical trials. While ZL-25 has been tested in mouse xenograft models and no overt toxicity was uncovered in their dosing, a full PK/PD profile should be generated, and work with medicinal chemists will help to refine the structure to find compounds that are more effective at lower doses. Also, the IC_{50} and affinity values for ZL-25 could certainly be better, and the suggested structures should be synthesized and tested not only against Skp2, but against other F-box proteins to ensure that specificity towards Skp2 is not lost.

For the ZM-276 inhibitor, it is certainly more potent than ZL-25, and it is more specific in targeting p27 to be restored in the nucleus, but at this point in its development, we don't have as concrete of evidence that it is precisely binding in the mode that has been predicted (not as concrete as ZL-25). Therefore, mutagenesis studies should be carried out to remedy this, as well as crystallographic studies.

Skp2 is an oncogene in the ubiquitin pathway and this finding validates Skp2 as a potential clinical target for which small-molecule inhibitors can be designed. These novel inhibitors are a first-in-class protein-protein interaction (PPI) inhibitors of the Skp1-Skp2 complex. Disruption of this complex halts function of the SCF complex and restores p21/p27. Preclinical work is ongoing to generate a more

efficient lead, and to identify a suitable cancer for development. This work also shows that targeting PPIs are a high-risk yet high-reward strategy that opens up a new domain of potential druggable targets. Computational techniques outlined here can help minimize this risk by using hot spot prediction to identify targetable regions on proteins.

As this study identifies two regions on which to target Skp2 with slightly different mechanisms of action (one disrupting the Skp1/Skp2 interaction, the other being disrupting the Cks1/Skp2/p27 interaction), it stands to reason that perhaps these two mechanisms and potential treatment might be synergistic when used as a combination, especially for cancers that show oncogenic addiction towards Skp2 and the SCF complex. Certainly, as bortezomib (Velcade) is currently being used for the treatment of multiple myeloma as a proteasome inhibitor, the inhibitors outlined here should be considered as a second-line of defense for cancers that are resistant to bortezomib. However, it should be noted that there is a possibility that resistance can be possible via a novel proteolytic processing pathway that is independent of ubiquitination and results in the abrogation of p27 function via the elimination of the critical cyclin-binding domain (314), and KPC (Kip1 ubiquitination-promoting complex, which consists of KPC1 and KPC2, a RING-finger domain and a ubiquitin-like domain, respectively) targets p27 for degradation in the G_0 - G_1 transition while Skp2/SCF complex targets it in the nucleus in the S and G_2 phases (238); so it is possible that a compensation via KPC might make both inhibitors ineffective. Further

studies should be carried out to examine if these effects could be problematic for the development of Skp2 inhibition.

Further work would ideally involve resolving the crystal structure of any of the Skp2/inhibitor complex (or complexes) which would give a highly affirmative confirmation of the exact binding pose of these inhibitors, as well as show how Skp2 and the F-box and Cks1 domains respond to binding. Additionally a resolved crystal structure of the Skp1/Skp2/ZL-25 complex as well as the complete SCF complex with ligands could indicate how ZL-25 is able to disrupt the preformed SCF complexes proteins. These structures could be used to conclusively suggest modifications for all inhibitors that would result in increased binding and efficacy. Additionally other means by which to measure inhibitor binding such ass ITC or Biacore studies to quantify the binding affinity and also be used for further design and QSAR studies.

In addition to resolving the binding affinity experimentally, performing advanced QM/MM molecular dynamics of the complex to attempt to calculate a binding affinity of the inhibitor. This binding energy would then be compared to a series of similar compounds to the original (as well as compared relative to any ITC/Biocore data generated of the protein-ligand complex) and be used as a basis for qualitative structure-activity relationship (QSAR) studies.

Use of new techniques in functional-genetic approaches can help answer questions about the best use and classification of these novel inhibitors. One approach is to examine the response of these inhibitors using a panel of tumor cell lines, and creating a signature of the sensitivity and resistance, similar to what is available in the

NCI-60 panel (315). Another approach is to analyze the global changes in the transcript that are inducted by a test ligand and compare that by known drugs or defined alterations in the genome (316-319). From this, common changes in expression are used to cluster small molecules that are similar. Finally, using an RNAi-based approach combined with machine learning technique can characterize therapeutic agents, especially in cancer (320). This allows the prediction of a mechanism of action based on the shRNA signature and can characterize a diverse range of the categories at high resolution.

One other aspect of this work that should be explored is the question as to how ZL-25 specifically attenuates the proliferation and survival of cancer cells. It was shown that ZL-25 has multiple effects on conditions such as: apoptosis, aerobic glycolysis, and cellular senescence in a p53-independent manner (176), but it is unclear which of these effects leads most directly to the decreased proliferation and survival of cancer cells, and if these effects are different depending on the tumor type.

From the NCI-60 data, it has been shown that leukemia cell lines seem to respond the highest to ZL-25. This would point to these types of cancers as potentially being a favorable indication to focus on. However, as the ZL-25-treated mouse xenograft data indicates, solid tumors also show good response rates. Identifying the precise cancer and cancer subtype are crucial towards maximizing the potential of ZL-25 and the other inhibitors targeting Skp2.

Appendix

1. Docking parameters

For the precise parameters used in GOLD 5.0.1 (CCDC) for the virtual screening and docking, the primary scoring function used for the virtual screen was GoldScore, whereas GoldScore and occasionally ChemScore as well as ASP scoring functions were used to rescore selected ligands for docking analysis and QSAR work. Also for the virtual screen the setting of "allow early termination" and “soft potentials” were turned off, and a search efficiency set to 200% was employed to allow maximal exploration of the hits. For the rescoring and QSAR work, these two settings were set in the on state.

Configuration file for HiPCDock (AutoDock) used for screening on Skp2/Skp1 interaction site on pocket1 (pocket 2 used grid_CenterCoordinates="-1.962 -100.846 10.708"):

[job]

[dockingSoftware]

dockingSoftwarePackage="AutoDock3"

[autoDockReceptorPreProcessing]

receptor="/scratch2/szhang/SKP2/AutoDock/skip2_2AST_with_H.pdb"

[autoGridParameters]

grid_Center="By Coordinates"

grid_CenterCoordinates="1.134 -114.378 -1.751"

grid_Points="80 80 80"

spacing="0.25"

[autoDockLigandPreProcessing]

database="/scratch2/lducuny/DataBase/ChemBridge/Chem_Lig"

processLigand="Yes"

autotorsFlags="-m -h -o -a -b -c -M -A +6.5"

[autoDockSearchingParameters]

searchingMethod="Genetic Algorithm"

geneticAlgorithmParameters

numberOfGARuns="100"

populationSize="100"

maximumNumberOfEnergyEvaluations="500000"

maximumNumberOfGenerations="27000"

maximumNumberOfTopIndividualsThatAutomaticallySurvive="1"

rateOfGeneMutation="0.02"

rateOfCrossover="0.8"

meanOfCauchyDistributionForGeneMutation="0.0"

varianceOfCauchyDistributionForGeneMutation="1.0"

numberOfGenerationsForPickingWorstIndividual="10"

energyParameters

externalGridEnergy="1000.0"

maximumAllowableInitialEnrgy="0.0"

maximumNumberOfRetries="10000"

stepSizeParameters

translation="2.0"

quaternion="50.0"

torsion="50.0"

outputFormatParameters

rmsClusterTolerance="0.5"

performAClusterAnalysis="Yes"

[outputOptions]

selectionCriterion1Name="Ranking"

selectionCriterion1="5"

selectionCriterion2Name="Binding Affinity"

selectionCriterion2="6"

outputFormat="Text File"

Bibliography

1. Overington, J. P., B. Al-Lazikani, and A. L. Hopkins. 2006. How many drug targets are there? *Nat Rev Drug Discov* 5: 993-996.
2. Santos, R., O. Ursu, A. Gaulton, A. P. Bento, R. S. Donadi, C. G. Bologa, A. Karlsson, B. Al-Lazikani, A. Hersey, T. I. Oprea, and J. P. Overington. 2017. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* 16: 19-34.
3. Imming, P., C. Sinning, and A. Meyer. 2006. Drugs, their targets and the nature and number of drug targets. *Nat Rev Drug Discov* 5: 821-834.
4. Golden, J. B. 2003. Prioritizing the human genome: knowledge management for drug discovery. *Curr Opin Drug Discov Devel* 6: 310-316.
5. Gaulton, A., A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrian-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magarinos, J. P. Overington, G. Papadatos, I. Smit, and A. R. Leach. 2017. The ChEMBL database in 2017. *Nucleic Acids Res* 45: D945-D954.
6. Varmus, H. 2006. The new era in cancer research. *Science* 312: 1162-1165.
7. Gavin, A. C., M. Bosche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A. M. Michon, C. M. Cruciat, M. Remor, C. Hofert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M. A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes,

M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer, and G. Superti-Furga. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141-147.

8. Johnson, J. E. 1996. Functional implications of protein-protein interactions in icosahedral viruses. *Proc Natl Acad Sci U S A* 93: 27-33.

9. Chothia, C., and J. Janin. 1975. Principles of protein-protein recognition. *Nature* 256: 705-708.

10. Lawrence, M. C., and P. M. Colman. 1993. Shape complementarity at protein/protein interfaces. *J Mol Biol* 234: 946-950.

11. Norel, R., S. L. Lin, H. J. Wolfson, and R. Nussinov. 1994. Shape complementarity at protein-protein interfaces. *Biopolymers* 34: 933-940.

12. Lo Conte, L., C. Chothia, and J. Janin. 1999. The atomic structure of protein-protein recognition sites. *J Mol Biol* 285: 2177-2198.

13. Keskin, O., B. Ma, and R. Nussinov. 2005. Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues. *J Mol Biol* 345: 1281-1294.

14. Janin, J. 1995. Elusive affinities. *Proteins* 21: 30-39.

15. Janin, J. 1995. Protein-protein recognition. *Prog Biophys Mol Biol* 64: 145-166.

16. Jones, S., and J. M. Thornton. 1996. Principles of protein-protein interactions. *Proc Natl Acad Sci U S A* 93: 13-20.

17. Janin, J., and C. Chothia. 1990. The structure of protein-protein recognition sites. *J Biol Chem* 265: 16027-16030.

18. Tsai, C. J., and R. Nussinov. 1997. Hydrophobic folding units at protein-protein interfaces: implications to protein folding and to protein-protein association. *Protein Sci* 6: 1426-1437.

19. Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci* 1: 169-181.

20. Wells, J. A., and C. L. McClendon. 2007. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature* 450: 1001-1009.

21. Verkhivker, G. M., D. Bouzida, D. K. Gehlhaar, P. A. Rejto, S. T. Freer, and P. W. Rose. 2003. Computational detection of the binding-site hot spot at the remodeled human growth hormone-receptor interface. *Proteins* 53: 201-219.

22. Aqvist, J., V. B. Luzhkov, and B. O. Brandsdal. 2002. Ligand binding affinities from MD simulations. *Acc Chem Res* 35: 358-365.

23. Marcotte, E. M., M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. 1999. Detecting protein function and protein-protein interactions from genome sequences. *Science* 285: 751-753.

24. Blazer, L. L., and R. R. Neubig. 2009. Small molecule protein-protein interaction inhibitors as CNS therapeutic agents: current progress and future hurdles. *Neuropsychopharmacology* 34: 126-141.

25. White, A. W., A. D. Westwell, and G. Brahemi. 2008. Protein-protein interactions as targets for small-molecule therapeutics in cancer. *Expert Rev Mol Med* 10: e8.

26. Ajay, A., W. P. Walters, and M. A. Murcko. 1998. Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? *J Med Chem* 41: 3314-3324.

27. Lipinski, C. A., F. Lombardo, B. W. Dominy, and P. J. Feeney. 2001. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 46: 3-26.

28. Siehler, S. 2008. Cell-based assays in GPCR drug discovery. *Biotechnol J* 3: 471-483.

29. Hopkins, A. L., and C. R. Groom. 2002. The druggable genome. *Nat Rev Drug Discov* 1: 727-730.

30. Ofran, Y., M. Punta, R. Schneider, and B. Rost. 2005. Beyond annotation transfer by homology: novel protein-function prediction methods to assist drug discovery. *Drug Discov Today* 10: 1475-1482.

31. Keskin, O., A. Gursoy, B. Ma, and R. Nussinov. 2007. Towards drugs targeting multiple proteins in a systems biology approach. *Curr Top Med Chem* 7: 943-951.

32. Sharma, S. K., T. M. Ramsey, and K. W. Bair. 2002. Protein-protein interactions: lessons learned. *Curr Med Chem Anticancer Agents* 2: 311-330.

33. Clackson, T., and J. A. Wells. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* 267: 383-386.

34. Bogan, A. A., and K. S. Thorn. 1998. Anatomy of hot spots in protein interfaces. *J Mol Biol* 280: 1-9.

35. Wells, J. A. 1991. Systematic mutational analyses of protein-protein interfaces. *Methods Enzymol* 202: 390-411.

36. Schreiber, G., and A. R. Fersht. 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J Mol Biol* 248: 478-486.

37. Stites, W. E. 1997. Proteinminus signProtein Interactions: Interface Structure, Binding Thermodynamics, and Mutational Analysis. *Chem Rev* 97: 1233-1250.

38. Clackson, T., M. H. Ultsch, J. A. Wells, and A. M. de Vos. 1998. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J Mol Biol* 277: 1111-1128.

39. Hu, Z., B. Ma, H. Wolfson, and R. Nussinov. 2000. Conservation of polar residues as hot spots at protein interfaces. *Proteins* 39: 331-342.

40. Kouadio, J. L., J. R. Horn, G. Pal, and A. A. Kossiakoff. 2005. Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface. *J Biol Chem* 280: 25524-25532.

41. Thorn, K. S., and A. A. Bogan. 2001. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* 17: 284-285.

42. Moreira, I. S., P. A. Fernandes, and M. J. Ramos. 2007. Hot spots--a review of the protein-protein interface determinant amino-acid residues. *Proteins* 68: 803-812.

43. Lichtarge, O., H. R. Bourne, and F. E. Cohen. 1996. An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* 257: 342-358.

44. Samanta, U., D. Pal, and P. Chakrabarti. 2000. Environment of tryptophan side chains in proteins. *Proteins* 38: 288-300.

45. Fernandez, A. 2002. Desolvation shell of hydrogen bonds in folded proteins, protein complexes and folding pathways. *FEBS Lett* 527: 166-170.

46. Tuncbag, N., O. Keskin, and A. Gursoy. 2010. HotPoint: hot spot prediction server for protein interfaces. *Nucleic Acids Res* 38: W402-406.

47. Li, X., O. Keskin, B. Ma, R. Nussinov, and J. Liang. 2004. Protein-protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. *J Mol Biol* 344: 781-795.

48. Ma, B., T. Elkayam, H. Wolfson, and R. Nussinov. 2003. Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A* 100: 5772-5777.

49. Lockless, S. W., and R. Ranganathan. 1999. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 286: 295-299.

50. Caffrey, D. R., S. Somaroo, J. D. Hughes, J. Mintseris, and E. S. Huang. 2004. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci* 13: 190-202.

51. Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. *Science* 296: 750-752.

52. Panchenko, A. R., F. Kondrashov, and S. Bryant. 2004. Prediction of functional sites by analysis of sequence and structure conservation. *Protein Sci* 13: 884-892.

53. DeLano, W. L. 2002. Unraveling hot spots in binding interfaces: progress and challenges. *Curr Opin Struct Biol* 12: 14-20.

54. Wells, J. A. 1996. Binding in the growth hormone receptor complex. *Proc Natl Acad Sci U S A* 93: 1-6.

55. Wells, J. A., and A. M. de Vos. 1993. Structure and function of human growth hormone: implications for the hematopoietins. *Annu Rev Biophys Biomol Struct* 22: 329-351.

56. Thanos, C. D., W. L. DeLano, and J. A. Wells. 2006. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci U S A* 103: 15422-15427.

57. Arkin, M. R., and J. A. Wells. 2004. Small-molecule inhibitors of protein-protein interactions: progressing towards the dream. *Nat Rev Drug Discov* 3: 301-317.

58. Wrighton, N. C., F. X. Farrell, R. Chang, A. K. Kashyap, F. P. Barbone, L. S. Mulcahy, D. L. Johnson, R. W. Barrett, L. K. Jolliffe, and W. J. Dower. 1996. Small peptides as potent mimetics of the protein hormone erythropoietin. *Science* 273: 458-464.

59. Sidhu, S. S., H. B. Lowman, B. C. Cunningham, and J. A. Wells. 2000. Phage display for selection of novel binding peptides. *Methods Enzymol* 328: 333-363.

60. Livnah, O., E. A. Stura, D. L. Johnson, S. A. Middleton, L. S. Mulcahy, N. C. Wrighton, W. J. Dower, L. K. Jolliffe, and I. A. Wilson. 1996. Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 A. *Science* 273: 464-471.

61. DeLano, W. L., M. H. Ultsch, A. M. de Vos, and J. A. Wells. 2000. Convergent solutions to binding at a protein-protein interface. *Science* 287: 1279-1283.

62. Li, R., V. Dowd, D. J. Stewart, S. J. Burton, and C. R. Lowe. 1998. Design, synthesis, and application of a protein A mimetic. *Nat Biotechnol* 16: 190-195.

63. Tilley, J. W., L. Chen, D. C. Fry, S. D. Emerson, G. D. Powers, D. Biondi, T. Varnell, R. Trilles, R. Guthrie, F. Mennona, G. Kaplan, R. A. LeMahieu, M. Carson, R. J. Han, C. M. Liu, R. Palermo, and G. Ju. 1997. Identification of a

small molecule inhibitor of the IL-2/IL-2R alpha receptor interaction which binds to IL-2. *Journal of the American Chemical Society* 119: 7589-7590.

64. Gonzalez-Ruiz, D., and H. Gohlke. 2006. Targeting protein-protein interactions with small molecules: challenges and perspectives for computational binding epitope detection and ligand finding. *Curr Med Chem* 13: 2607-2625.

65. Rajamani, D., S. Thiel, S. Vajda, and C. J. Camacho. 2004. Anchor residues in protein-protein interactions. *Proc Natl Acad Sci U S A* 101: 11287-11292.

66. Acuner Ozbabacan, S. E., A. Gursoy, O. Keskin, and R. Nussinov. 2010. Conformational ensembles, signal transduction and residue hot spots: application to drug discovery. *Curr Opin Drug Discov Devel* 13: 527-537.

67. Wells, J. A. 1990. Additivity of mutational effects in proteins. *Biochemistry* 29: 8509-8517.

68. Cunningham, B. C., and J. A. Wells. 1989. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 244: 1081-1085.

69. Skolnick, J., J. S. Fetrow, and A. Kolinski. 2000. Structural genomics and its importance for gene function analysis. *Nat Biotechnol* 18: 283-287.

70. Morrison, K. L., and G. A. Weiss. 2001. Combinatorial alanine-scanning. *Curr Opin Chem Biol* 5: 302-307.

71. Dinner, A. R., A. Sali, L. J. Smith, C. M. Dobson, and M. Karplus. 2000. Understanding protein folding via free-energy surfaces from theory and experiment. *Trends Biochem Sci* 25: 331-339.

72. Kumar, S., B. Ma, C. J. Tsai, N. Sinha, and R. Nussinov. 2000. Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci* 9: 10-19.

73. Fischer, T. B., K. V. Arunachalam, D. Bailey, V. Mangual, S. Bakhru, R. Russo, D. Huang, M. Paczkowski, V. Lalchandani, C. Ramachandra, B. Ellison, S. Galer, J. Shapley, E. Fuentes, and J. Tsai. 2003. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. *Bioinformatics* 19: 1453-1454.

74. Piehler, J., and G. Schreiber. 2001. Fast transient cytokine-receptor interactions monitored in real time by reflectometric interference spectroscopy. *Anal Biochem* 289: 173-186.

75. Weiss, G. A., C. K. Watanabe, A. Zhong, A. Goddard, and S. S. Sidhu. 2000. Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci U S A* 97: 8950-8954.

76. Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano. 2005. The FoldX web server: an online force field. *Nucleic Acids Res* 33: W382-388.

77. Kim, D. E., D. Chivian, and D. Baker. 2004. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res* 32: W526-531.

78. Gao, Y., R. Wang, and L. Lai. 2004. Structure-based method for analyzing protein-protein interfaces. *J Mol Model* 10: 44-54.

79. Brenke, R., D. Kozakov, G. Y. Chuang, D. Beglov, D. Hall, M. R. Landon, C. Mattos, and S. Vajda. 2009. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* 25: 621-627.

80. Kozakov, D., L. E. Grove, D. R. Hall, T. Bohnuud, S. E. Mottarella, L. Luo, B. Xia, D. Beglov, and S. Vajda. 2015. The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat Protoc* 10: 733-755.

81. Assi, S. A., T. Tanaka, T. H. Rabbitts, and N. Fernandez-Fuentes. 2010. PCRPi: Presaging Critical Residues in Protein interfaces, a new computational tool to chart hot spots in protein interfaces. *Nucleic Acids Res* 38: e86.

82. Segura Mora, J., S. A. Assi, and N. Fernandez-Fuentes. 2010. Presaging critical residues in protein interfaces-web server (PCRPi-W): a web server to chart hot spots in protein interfaces. *PLoS One* 5: e12352.

83. Guharoy, M., and P. Chakrabarti. 2009. Empirical estimation of the energetic contribution of individual interface residues in structures of protein-protein complexes. *J Comput Aided Mol Des* 23: 645-654.

84. Zhu, X., and J. C. Mitchell. 2011. KFC2: A knowledge-based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins* 79: 2671-2683.

85. Cho, K. I., D. Kim, and D. Lee. 2009. A feature-based approach to modeling protein-protein interaction hot spots. *Nucleic Acids Res* 37: 2672-2687.

86. Lise, S., C. Archambeau, M. Pontil, and D. T. Jones. 2009. Prediction of hot spot residues at protein-protein interfaces by combining machine learning and energy-based methods. *BMC Bioinformatics* 10: 365.

87. Lise, S., D. Buchan, M. Pontil, and D. T. Jones. 2011. Predictions of hot spot residues at protein-protein interfaces using support vector machines. *PLoS One* 6: e16774.

88. Higa, R. H., and C. L. Tozzi. 2009. Prediction of binding hot spot residues by using structural and evolutionary parameters. *Genet Mol Biol* 32: 626-633.

89. Grosdidier, S., and J. Fernandez-Recio. 2008. Identification of hot-spot residues in protein-protein interactions by computational docking. *BMC Bioinformatics* 9: 447.

90. Guney, E., N. Tuncbag, O. Keskin, and A. Gursoy. 2008. HotSprint: database of computational hot spots in protein interfaces. *Nucleic Acids Res* 36: D662-666.

91. Tuncbag, N., G. Kar, O. Keskin, A. Gursoy, and R. Nussinov. 2009. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Brief Bioinform* 10: 217-232.

92. Guerois, R., J. E. Nielsen, and L. Serrano. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320: 369-387.

93. Kortemme, T., D. E. Kim, and D. Baker. 2004. Computational alanine scanning of protein-protein interfaces. *Sci STKE* 2004: pl2.

94. Kortemme, T., and D. Baker. 2002. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* 99: 14116-14121.

95. Chen, J., and L. Lai. 2006. Pocket v.2: further developments on receptor-based pharmacophore modeling. *J Chem Inf Model* 46: 2684-2691.

96. Mattos, C., and D. Ringe. 1996. Locating and characterizing binding sites on proteins. *Nat Biotechnol* 14: 595-599.

97. Tuncbag, N., A. Gursoy, and O. Keskin. 2009. Identification of computational hot spots in protein interfaces: combining solvent accessibility and inter-residue potentials improves the accuracy. *Bioinformatics* 25: 1513-1520.

98. Berman, H. M., T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J. D. Westbrook, and C. Zardecki. 2002. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr* 58: 899-907.

99. Segura, J., and N. Fernandez-Fuentes. 2011. PCRPi-DB: a database of computationally annotated hot spots in protein interfaces. *Nucleic Acids Res* 39: D755-760.

100. Yogurtcu, O. N., S. B. Erdemli, R. Nussinov, M. Turkay, and O. Keskin. 2008. Restricted mobility of conserved residues in protein-protein interfaces in molecular simulations. *Biophys J* 94: 3475-3485.

101. Kimura, S. R., R. C. Brower, S. Vajda, and C. J. Camacho. 2001. Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys J* 80: 635-642.

102. Huo, S., I. Massova, and P. A. Kollman. 2002. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J Comput Chem* 23: 15-27.

103. Massova, I., and P. A. Kollman. 1999. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem Soc.* 121: 8133-8143.

104. Mitchell, J. C., R. Kerr, and L. F. Ten Eyck. 2001. Rapid atomic density methods for molecular shape characterization. *J Mol Graph Model* 19: 325-330, 388-390.

105. Darnell, S. J., L. LeGault, and J. C. Mitchell. 2008. KFC Server: interactive forecasting of protein interaction hot spots. *Nucleic Acids Res* 36: W265-269.

106. Darnell, S. J., D. Page, and J. C. Mitchell. 2007. An automated decision-tree approach to predicting protein interaction hot spots. *Proteins* 68: 813-823.

107. Wang, Y. T., Z. Y. Su, and C. L. Chen. 2009. Potential of mean force of the hepatitis C virus core protein-monoclonal 19D9D6 antibody interaction. *Biophys Chem* 145: 86-90.

108. Noble, W. S. 2006. What is a support vector machine? *Nat Biotechnol* 24: 1565-1567.

109. Halperin, I., H. Wolfson, and R. Nussinov. 2004. Protein-protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. *Structure* 12: 1027-1038.

110. Fernandez-Recio, J., M. Totrov, and R. Abagyan. 2004. Identification of protein-protein interaction sites from docking energy landscapes. *J Mol Biol* 335: 843-865.

111. Keskin, O., I. Bahar, A. Y. Badretdinov, O. B. Ptitsyn, and R. L. Jernigan. 1998. Empirical solvent-mediated potentials hold for both intra-molecular and inter-molecular inter-residue interactions. *Protein Sci* 7: 2578-2586.

112. Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature* 319: 199-203.

113. Hershko, A., and A. Ciechanover. 1998. The ubiquitin system. *Annu Rev Biochem* 67: 425-479.

114. Hershko, A. 1983. Ubiquitin: roles in protein modification and breakdown. *Cell* 34: 11-12.

115. Handley, P. M., M. Mueckler, N. R. Siegel, A. Ciechanover, and A. L. Schwartz. 1991. Molecular cloning, sequence, and tissue distribution of the human ubiquitin-activating enzyme E1. *Proc Natl Acad Sci U S A* 88: 258-262.

116. Schulman, B. A., and J. W. Harper. 2009. Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* 10: 319-331.

117. Jentsch, S. 1992. The ubiquitin-conjugation system. *Annual review of genetics* 26: 179-207.

118. Berndsen, C. E., and C. Wolberger. 2014. New insights into ubiquitin E3 ligase mechanism. *Nat Struct Mol Biol* 21: 301-307.

119. Nakayama, K. I., and K. Nakayama. 2006. Ubiquitin ligases: cell-cycle control and cancer. *Nat Rev Cancer* 6: 369-381.

120. Hershko, A., and A. Ciechanover. 1992. The ubiquitin system for protein degradation. *Annu Rev Biochem* 61: 761-807.

121. Ciechanover, A. 1994. The ubiquitin-proteasome proteolytic pathway. *Cell* 79: 13-21.

122. Rieser, E., S. M. Cordier, and H. Walczak. 2013. Linear ubiquitination: a newly discovered regulator of cell signalling. *Trends Biochem Sci* 38: 94-102.

123. Komander, D., and M. Rape. 2012. The ubiquitin code. *Annu Rev Biochem* 81: 203-229.

124. Chen, Z. J., and L. J. Sun. 2009. Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell* 33: 275-286.

125. Yang, W. L., X. Zhang, and H. K. Lin. 2010. Emerging role of Lys-63 ubiquitination in protein kinase and phosphatase activation and cancer development. *Oncogene* 29: 4493-4503.

126. Bach, I., and H. P. Ostendorff. 2003. Orchestrating nuclear functions: ubiquitin sets the rhythm. *Trends Biochem Sci* 28: 189-195.

127. Murray, A. W. 2004. Recycling the cell cycle: cyclins revisited. *Cell* 116: 221-234.

128. Weissman, A. M. 2001. Themes and variations on ubiquitylation. *Nat Rev Mol Cell Biol* 2: 169-178.

129. Pickart, C. M. 2004. Back to the future with ubiquitin. *Cell* 116: 181-190.

130. Glotzer, M., A. W. Murray, and M. W. Kirschner. 1991. Cyclin is degraded by the ubiquitin pathway. *Nature* 349: 132-138.

131. Farhana, L., M. Dawson, A. K. Rishi, Y. Zhang, E. Van Buren, C. Trivedi, U. Reichert, G. Fang, M. W. Kirschner, and J. A. Fontana. 2002. Cyclin B and E2F-1 expression in prostate carcinoma cells treated with the novel retinoid CD437 are regulated by the ubiquitin-mediated pathway. *Cancer research* 62: 3842-3849.

132. Bhattacharya, S., J. Garriga, J. Calbo, T. Yong, D. S. Haines, and X. Grana. 2003. SKP2 associates with p130 and accelerates p130 ubiquitylation and degradation in human cells. *Oncogene* 22: 2443-2451.

133. Hoeller, D., and I. Dikic. 2009. Targeting the ubiquitin system in cancer therapy. *Nature* 458: 438-444.

134. Lipkowitz, S., and A. M. Weissman. 2011. RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. *Nat Rev Cancer* 11: 629-643.

135. Kirkin, V., and I. Dikic. 2011. Ubiquitin networks in cancer. *Curr Opin Genet Dev* 21: 21-28.

136. Pal, A., M. A. Young, and N. J. Donato. 2014. Emerging potential of therapeutic targeting of ubiquitin-specific proteases in the treatment of cancer. *Cancer Res* 74: 4955-4966.

137. Miranda, M., and A. Sorkin. 2007. Regulation of receptors and transporters by ubiquitination: new insights into surprisingly similar mechanisms. *Mol Interv* 7: 157-167.

138. Hicke, L., H. L. Schubert, and C. P. Hill. 2005. Ubiquitin-binding domains. *Nat Rev Mol Cell Biol* 6: 610-621.

139. Pickart, C. M. 2001. Ubiquitin enters the new millennium. *Mol Cell* 8: 499-504.

140. Sigismund, S., S. Polo, and P. P. Di Fiore. 2004. Signaling through monoubiquitination. *Curr Top Microbiol Immunol* 286: 149-185.

141. Hicke, L. 2001. Protein regulation by monoubiquitin. *Nat Rev Mol Cell Biol* 2: 195-201.

142. Adams, J., and M. Kauffman. 2004. Development of the proteasome inhibitor Velcade (Bortezomib). *Cancer Invest* 22: 304-311.

143. Frankland-Searby, S., and S. R. Bhaumik. 2012. The 26S proteasome complex: an attractive target for cancer therapy. *Biochim Biophys Acta* 1825: 64-76.

144. Kumar, S. K., S. V. Rajkumar, A. Dispenzieri, M. Q. Lacy, S. R. Hayman, F. K. Buadi, S. R. Zeldenrust, D. Dingli, S. J. Russell, J. A. Lust, P. R. Greipp, R. A. Kyle, and M. A. Gertz. 2008. Improved survival in multiple myeloma and the impact of novel therapies. *Blood* 111: 2516-2520.

145. Bonvini, P., E. Zorzi, G. Basso, and A. Rosolen. 2007. Bortezomib-mediated 26S proteasome inhibition causes cell-cycle arrest and induces apoptosis in CD-30+ anaplastic large cell lymphoma. *Leukemia* 21: 838-842.

146. Groll, M., C. R. Berkers, H. L. Ploegh, and H. Ovaa. 2006. Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20S proteasome. *Structure* 14: 451-456.

147. Hideshima, T., P. G. Richardson, and K. C. Anderson. 2011. Mechanism of action of proteasome inhibitors and deacetylase inhibitors and the biological basis of synergy in multiple myeloma. *Mol Cancer Ther* 10: 2034-2042.

148. Gelman, J. S., J. Sironi, I. Berezniuk, S. Dasgupta, L. M. Castro, F. C. Gozzo, E. S. Ferro, and L. D. Fricker. 2013. Alterations of the intracellular peptidome in response to the proteasome inhibitor bortezomib. *PLoS One* 8: e53263.

149. Mattern, M. R., J. Wu, and B. Nicholson. 2012. Ubiquitin-based anticancer therapy: carpet bombing with proteasome inhibitors vs surgical strikes with E1, E2, E3, or DUB inhibitors. *Biochim Biophys Acta* 1823: 2014-2021.

150. Kubiczkova, L., L. Pour, L. Sedlarikova, R. Hajek, and S. Sevcikova. 2014. Proteasome inhibitors - molecular basis and current perspectives in multiple myeloma. *J Cell Mol Med* 18: 947-961.

151. Richardson, P. G., H. Briemberg, S. Jagannath, P. Y. Wen, B. Barlogie, J. Berenson, S. Singhal, D. S. Siegel, D. Irwin, M. Schuster, G. Srkalovic, R. Alexanian, S. V. Rajkumar, S. Limentani, M. Alsina, R. Z. Orlowski, K. Najarian, D. Esseltine, K. C. Anderson, and A. A. Amato. 2006. Frequency, characteristics, and reversibility of peripheral neuropathy during treatment of advanced multiple myeloma with bortezomib. *J Clin Oncol* 24: 3113-3120.

152. Richardson, P. G., B. Barlogie, J. Berenson, S. Singhal, S. Jagannath, D. Irwin, S. V. Rajkumar, G. Srkalovic, M. Alsina, R. Alexanian, D. Siegel, R. Z. Orlowski, D. Kuter, S. A. Limentani, S. Lee, T. Hideshima, D. L. Esseltine, M. Kauffman, J. Adams, D. P. Schenkein, and K. C. Anderson. 2003. A phase 2 study of bortezomib in relapsed, refractory myeloma. *N Engl J Med* 348: 2609-2617.

153. Suzuki, E., S. Demo, E. Deu, J. Keats, S. Arastu-Kapur, P. L. Bergsagel, M. K. Bennett, and C. J. Kirk. 2011. Molecular mechanisms of bortezomib resistant adenocarcinoma cells. *PLoS One* 6: e27996.

154. Vij, R., D. S. Siegel, S. Jagannath, A. J. Jakubowiak, A. K. Stewart, K. McDonagh, N. Bahlis, A. Belch, L. A. Kunkel, S. Wear, A. F. Wong, and M. Wang. 2012. An open-label, single-arm, phase 2 study of single-agent carfilzomib in patients with relapsed and/or refractory multiple myeloma who have been previously treated with bortezomib. *Br J Haematol* 158: 739-748.

155. Vij, R., M. Wang, J. L. Kaufman, S. Lonial, A. J. Jakubowiak, A. K. Stewart, V. Kukreti, S. Jagannath, K. T. McDonagh, M. Alsina, N. J. Bahlis, F. J. Reu,

N. Y. Gabrail, A. Belch, J. V. Matous, P. Lee, P. Rosen, M. Sebag, D. H. Vesole, L. A. Kunkel, S. M. Wear, A. F. Wong, R. Z. Orlowski, and D. S. Siegel. 2012. An open-label, single-arm, phase 2 (PX-171-004) study of single-agent carfilzomib in bortezomib-naive patients with relapsed and/or refractory multiple myeloma. *Blood* 119: 5661-5670.

156. Meng, L., R. Mohan, B. H. Kwok, M. Elofsson, N. Sin, and C. M. Crews. 1999. Epoxomicin, a potent and selective proteasome inhibitor, exhibits in vivo antiinflammatory activity. *Proc Natl Acad Sci U S A* 96: 10403-10408.

157. Potts, B. C., M. X. Albitar, K. C. Anderson, S. Baritaki, C. Berkers, B. Bonavida, J. Chandra, D. Chauhan, J. C. Cusack, Jr., W. Fenical, I. M. Ghobrial, M. Groll, P. R. Jensen, K. S. Lam, G. K. Lloyd, W. McBride, D. J. McConkey, C. P. Miller, S. T. Neuteboom, Y. Oki, H. Ovaa, F. Pajonk, P. G. Richardson, A. M. Roccaro, C. M. Sloss, M. A. Spear, E. Valashi, A. Younes, and M. A. Palladino. 2011. Marizomib, a proteasome inhibitor for all seasons: preclinical profile and a framework for clinical trials. *Curr Cancer Drug Targets* 11: 254-284.

158. Kumar, S. K., J. G. Berdeja, R. Niesvizky, S. Lonial, J. P. Laubach, M. Hamadani, A. K. Stewart, P. Hari, V. Roy, R. Vescio, J. L. Kaufman, D. Berg, E. Liao, A. Di Bacco, J. Estevam, N. Gupta, A. M. Hui, V. Rajkumar, and P. G. Richardson. 2014. Safety and tolerability of ixazomib, an oral proteasome inhibitor, in combination with lenalidomide and dexamethasone in patients

with previously untreated multiple myeloma: an open-label phase 1/2 study. *Lancet Oncol* 15: 1503-1512.

159. Piva, R., B. Ruggeri, M. Williams, G. Costa, I. Tamagno, D. Ferrero, V. Giai, M. Coscia, S. Peola, M. Massaia, G. Pezzoni, C. Allievi, N. Pescalli, M. Cassin, S. di Giovine, P. Nicoli, P. de Feudis, I. Strepponi, I. Roato, R. Ferracini, B. Bussolati, G. Camussi, S. Jones-Bolin, K. Hunter, H. Zhao, A. Neri, A. Palumbo, C. Berkers, H. Ovaa, A. Bernareggi, and G. Inghirami. 2008. CEP-18770: A novel, orally active proteasome inhibitor with a tumor-selective pharmacologic profile competitive with bortezomib. *Blood* 111: 2765-2775.

160. Tsukamoto, S., T. Takeuchi, H. Rotinsulu, R. E. Mangindaan, R. W. van Soest, K. Ukai, H. Kobayashi, M. Namikoshi, T. Ohta, and H. Yokosawa. 2008. Leucettamol A: a new inhibitor of Ubc13-Uev1A interaction isolated from a marine sponge, Leucetta aff. microrhaphis. *Bioorg Med Chem Lett* 18: 6319-6320.

161. Dalisay, D. S., S. Tsukamoto, and T. F. Molinski. 2009. Absolute configuration of the alpha,omega-bifunctionalized sphingolipid leucettamol A from Leucetta microrhaphis by deconvoluted exciton coupled CD. *J Nat Prod* 72: 353-359.

162. Ceccarelli, D. F., X. Tang, B. Pelletier, S. Orlicky, W. Xie, V. Plantevin, D. Neculai, Y. C. Chou, A. Ogunjimi, A. Al-Hakim, X. Varelas, J. Koszela, G. A. Wasney, M. Vedadi, S. Dhe-Paganon, S. Cox, S. Xu, A. Lopez-Girona, F.

Mercurio, J. Wrana, D. Durocher, S. Meloche, D. R. Webb, M. Tyers, and F. Sicheri. 2011. An allosteric inhibitor of the human Cdc34 ubiquitin-conjugating enzyme. *Cell* 145: 1075-1087.

163. Ito, T., H. Ando, T. Suzuki, T. Ogura, K. Hotta, Y. Imamura, Y. Yamaguchi, and H. Handa. 2010. Identification of a primary target of thalidomide teratogenicity. *Science* 327: 1345-1350.

164. Kim, Y., and I. G. Schmidt-Wolf. 2015. Lenalidomide in multiple myeloma. *Expert Rev Anticancer Ther*: 1-7.

165. Lacy, M. Q., and S. V. Rajkumar. 2010. Pomalidomide: a new IMiD with remarkable activity in both multiple myeloma and myelofibrosis. *Am J Hematol* 85: 95-96.

166. Vassilev, L. T., B. T. Vu, B. Graves, D. Carvajal, F. Podlaski, Z. Filipovic, N. Kong, U. Kammlott, C. Lukacs, C. Klein, N. Fotouhi, and E. A. Liu. 2004. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303: 844-848.

167. Vassilev, L. T. 2007. MDM2 inhibitors for cancer therapy. *Trends Mol Med* 13: 23-31.

168. Vu, B., P. Wovkulich, G. Pizzolato, A. Lovey, Q. Ding, N. Jiang, J. J. Liu, C. Zhao, K. Glenn, Y. Wen, C. Tovar, K. Packman, L. Vassilev, and B. Graves. 2013. Discovery of RG7112: A Small-Molecule MDM2 Inhibitor in Clinical Development. *ACS Med Chem Lett* 4: 466-469.

169. Chang, Y. S., B. Graves, V. Guerlavais, C. Tovar, K. Packman, K. H. To, K. A. Olson, K. Kesavan, P. Gangurde, A. Mukherjee, T. Baker, K. Darlak, C. Elkin, Z. Filipovic, F. Z. Qureshi, H. Cai, P. Berry, E. Feyfant, X. E. Shi, J. Horstick, D. A. Annis, A. M. Manning, N. Fotouhi, H. Nash, L. T. Vassilev, and T. K. Sawyer. 2013. Stapled alpha-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc Natl Acad Sci U S A* 110: E3445-3454.

170. Gembarska, A., F. Luciani, C. Fedele, E. A. Russell, M. Dewaele, S. Villar, A. Zwolinska, S. Haupt, J. de Lange, D. Yip, J. Goydos, J. J. Haigh, Y. Haupt, L. Larue, A. Jochemsen, H. Shi, G. Moriceau, R. S. Lo, G. Ghanem, M. Shackleton, F. Bernal, and J. C. Marine. 2012. MDM4 is a key therapeutic target in cutaneous melanoma. *Nat Med* 18: 1239-1247.

171. Wang, H., X. Ma, S. Ren, J. K. Buolamwini, and C. Yan. 2011. A small-molecule inhibitor of MDMX activates p53 and induces apoptosis. *Mol Cancer Ther* 10: 69-79.

172. Ungermannova, D., J. Lee, G. Zhang, H. G. Dallmann, C. S. McHenry, and X. Liu. 2013. High-throughput screening AlphaScreen assay for identification of small-molecule inhibitors of ubiquitin E3 ligase SCFSkp2-Cks1. *J Biomol Screen* 18: 910-920.

173. Orlicky, S., X. Tang, V. Neduva, N. Elowe, E. D. Brown, F. Sicheri, and M. Tyers. 2010. An allosteric inhibitor of substrate recognition by the SCF(Cdc4) ubiquitin ligase. *Nat Biotechnol* 28: 733-737.

174. Wu, L., A. V. Grigoryan, Y. Li, B. Hao, M. Pagano, and T. J. Cardozo. 2012. Specific small molecule inhibitors of Skp2-mediated p27 degradation. *Chem Biol* 19: 1515-1524.

175. Chen, Q., W. Xie, D. J. Kuhn, P. M. Voorhees, A. Lopez-Girona, D. Mendy, L. G. Corral, V. P. Krenitsky, W. Xu, L. Moutouh-de Parseval, D. R. Webb, F. Mercurio, K. I. Nakayama, K. Nakayama, and R. Z. Orlowski. 2008. Targeting the p27 E3 ligase SCF(Skp2) results in p27- and Skp2-mediated cell-cycle arrest and activation of autophagy. *Blood* 111: 4690-4699.

176. Chan, C. H., J. K. Morrow, C. F. Li, Y. Gao, G. Jin, A. Moten, L. J. Stagg, J. E. Ladbury, Z. Cai, D. Xu, C. J. Logothetis, M. C. Hung, S. Zhang, and H. K. Lin. 2013. Pharmacological inactivation of Skp2 SCF ubiquitin ligase restricts cancer stem cell traits and cancer progression. *Cell* 154: 556-568.

177. Nakajima, H., H. Fujiwara, Y. Furuichi, K. Tanaka, and N. Shimbara. 2008. A novel small-molecule inhibitor of NF-kappaB signaling. *Biochem Biophys Res Commun* 368: 1007-1013.

178. Blees, J. S., H. R. Bokesch, D. Rubsamen, K. Schulz, L. Milke, M. M. Bajer, K. R. Gustafson, C. J. Henrich, J. B. McMahon, N. H. Colburn, T. Schmid, and B. Brune. 2012. Erioflorin stabilizes the tumor suppressor Pdcd4 by inhibiting its interaction with the E3-ligase beta-TrCP1. *PLoS One* 7: e46567.

179. Chauhan, D., L. Catley, G. Li, K. Podar, T. Hideshima, M. Velankar, C. Mitsiades, N. Mitsiades, H. Yasui, A. Letai, H. Ovaa, C. Berkers, B. Nicholson, T. H. Chao, S. T. Neuteboom, P. Richardson, M. A. Palladino, and

K. C. Anderson. 2005. A novel orally active proteasome inhibitor induces apoptosis in multiple myeloma cells with mechanisms distinct from Bortezomib. *Cancer Cell* 8: 407-419.

180. Millward, M., T. Price, A. Townsend, C. Sweeney, A. Spencer, S. Sukumaran, A. Longenecker, L. Lee, A. Lay, G. Sharma, R. M. Gemmill, H. A. Drabkin, G. K. Lloyd, S. T. Neuteboom, D. J. McConkey, M. A. Palladino, and M. A. Spear. 2012. Phase 1 clinical trial of the novel proteasome inhibitor marizomib with the histone deacetylase inhibitor vorinostat in patients with melanoma, pancreatic and lung cancer based on in vitro assessments of the combination. *Invest New Drugs* 30: 2303-2317.

181. Dick, L. R., and P. E. Fleming. 2010. Building on bortezomib: second-generation proteasome inhibitors as anti-cancer therapy. *Drug Discov Today* 15: 243-249.

182. Kupperman, E., E. C. Lee, Y. Cao, B. Bannerman, M. Fitzgerald, A. Berger, J. Yu, Y. Yang, P. Hales, F. Bruzzese, J. Liu, J. Blank, K. Garcia, C. Tsu, L. Dick, P. Fleming, L. Yu, M. Manfredi, M. Rolfe, and J. Bolen. 2010. Evaluation of the proteasome inhibitor MLN9708 in preclinical models of human cancer. *Cancer Res* 70: 1970-1980.

183. Ushiyama, S., H. Umaoka, H. Kato, Y. Suwa, H. Morioka, H. Rotinsulu, F. Losung, R. E. Mangindaan, N. J. de Voogd, H. Yokosawa, and S. Tsukamoto. 2012. Manadosterols A and B, sulfonated sterol dimers inhibiting the Ubc13-

Uev1A interaction, isolated from the marine sponge Lissodendryx fibrosa. *J Nat Prod* 75: 1495-1499.

184. Metzger, M. B., V. A. Hristova, and A. M. Weissman. 2012. HECT and RING finger families of E3 ubiquitin ligases at a glance. *J Cell Sci* 125: 531-537.

185. Nijman, S. M., M. P. Luna-Vargas, A. Velds, T. R. Brummelkamp, A. M. Dirac, T. K. Sixma, and R. Bernards. 2005. A genomic and functional inventory of deubiquitinating enzymes. *Cell* 123: 773-786.

186. Skaar, J. R., J. K. Pagan, and M. Pagano. 2013. Mechanisms and function of substrate recruitment by F-box proteins. *Nat Rev Mol Cell Biol* 14: 369-381.

187. Lu, G., R. E. Middleton, H. Sun, M. Naniong, C. J. Ott, C. S. Mitsiades, K. K. Wong, J. E. Bradner, and W. G. Kaelin, Jr. 2014. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* 343: 305-309.

188. Fischer, E. S., K. Bohm, J. R. Lydeard, H. Yang, M. B. Stadler, S. Cavadini, J. Nagel, F. Serluca, V. Acker, G. M. Lingaraju, R. B. Tichkule, M. Schebesta, W. C. Forrester, M. Schirle, U. Hassiepen, J. Ottl, M. Hild, R. E. Beckwith, J. W. Harper, J. L. Jenkins, and N. H. Thoma. 2014. Structure of the DDB1-CRBN E3 ubiquitin ligase in complex with thalidomide. *Nature* 512: 49-53.

189. Lopez-Girona, A., D. Mendy, T. Ito, K. Miller, A. K. Gandhi, J. Kang, S. Karasawa, G. Carmel, P. Jackson, M. Abbasian, A. Mahmoudi, B. Cathers, E. Rychak, S. Gaidarova, R. Chen, P. H. Schafer, H. Handa, T. O. Daniel, J. F. Evans, and R. Chopra. 2012. Cereblon is a direct protein target for

immunomodulatory and antiproliferative activities of lenalidomide and pomalidomide. *Leukemia* 26: 2326-2335.

190. Zhu, Y. X., E. Braggio, C. X. Shi, L. A. Bruins, J. E. Schmidt, S. Van Wier, X. B. Chang, C. C. Bjorklund, R. Fonseca, P. L. Bergsagel, R. Z. Orlowski, and A. K. Stewart. 2011. Cereblon expression is required for the antimyeloma activity of lenalidomide and pomalidomide. *Blood* 118: 4771-4779.

191. Lee, J. T., and W. Gu. 2010. The multiple levels of regulation by p53 ubiquitination. *Cell Death Differ* 17: 86-92.

192. Haupt, Y., R. Maya, A. Kazaz, and M. Oren. 1997. Mdm2 promotes the rapid degradation of p53. *Nature* 387: 296-299.

193. Honda, R., H. Tanaka, and H. Yasuda. 1997. Oncoprotein MDM2 is a ubiquitin ligase E3 for tumor suppressor p53. *FEBS Lett* 420: 25-27.

194. Fang, S., J. P. Jensen, R. L. Ludwig, K. H. Vousden, and A. M. Weissman. 2000. Mdm2 is a RING finger-dependent ubiquitin protein ligase for itself and p53. *J Biol Chem* 275: 8945-8951.

195. Tabernero, J., L. Dirix, P. Schoffski, A. Cervantes, J. A. Lopez-Martin, J. Capdevila, L. van Beijsterveldt, S. Platero, B. Hall, Z. Yuan, R. Knoblauch, and S. H. Zhuang. 2011. A phase I first-in-human pharmacokinetic and pharmacodynamic study of serdemetan in patients with advanced solid tumors. *Clin Cancer Res* 17: 6313-6321.

196. Chargari, C., C. Leteur, E. Angevin, T. Bashir, B. Schoentjes, J. Arts, M. Janicot, J. Bourhis, and E. Deutsch. 2011. Preclinical assessment of JNJ-

26854165 (Serdemetan), a novel tryptamine compound with radiosensitizing activity in vitro and in tumor xenografts. *Cancer Lett* 312: 209-218.

197. Khoury, K., and A. Domling. 2012. P53 mdm2 inhibitors. *Curr Pharm Des* 18: 4668-4678.

198. Van Maerken, T., A. Rihani, A. Van Goethem, A. De Paepe, F. Speleman, and J. Vandesompele. 2014. Pharmacologic activation of wild-type p53 by nutlin therapy in childhood cancer. *Cancer Lett* 344: 157-165.

199. Joseph, T. L., A. Madhumalar, C. J. Brown, D. P. Lane, and C. S. Verma. 2010. Differential binding of p53 and nutlin to MDM2 and MDMX: computational studies. *Cell Cycle* 9: 1167-1181.

200. Deshaies, R. J., and C. A. Joazeiro. 2009. RING domain E3 ubiquitin ligases. *Annu Rev Biochem* 78: 399-434.

201. Petroski, M. D., and R. J. Deshaies. 2005. Function and regulation of cullin-RING ubiquitin ligases. *Nat Rev Mol Cell Biol* 6: 9-20.

202. Bielskiene, K., L. Bagdoniene, J. Mozuraitiene, B. Kazbariene, and E. Janulionis. 2015. E3 ubiquitin ligases as drug targets and prognostic biomarkers in melanoma. *Medicina (Kaunas)* 51: 1-9.

203. Aghajan, M., N. Jonai, K. Flick, F. Fu, M. Luo, X. Cai, I. Ouni, N. Pierce, X. Tang, B. Lomenick, R. Damoiseaux, R. Hao, P. M. Del Moral, R. Verma, Y. Li, C. Li, K. N. Houk, M. E. Jung, N. Zheng, L. Huang, R. J. Deshaies, P. Kaiser, and J. Huang. 2010. Chemical genetics screen for enhancers of

rapamycin identifies a specific inhibitor of an SCF family E3 ubiquitin ligase. *Nat Biotechnol* 28: 738-742.

204. Mallampalli, R. K., T. A. Coon, J. R. Glasser, C. Wang, S. R. Dunn, N. M. Weathington, J. Zhao, C. Zou, Y. Zhao, and B. B. Chen. 2013. Targeting F box protein Fbxo3 to control cytokine-driven inflammation. *J Immunol* 191: 5247-5255.

205. Chen, B. B., T. A. Coon, J. R. Glasser, B. J. McVerry, J. Zhao, Y. Zhao, C. Zou, B. Ellis, F. C. Sciurba, Y. Zhang, and R. K. Mallampalli. 2013. A combinatorial F box protein directed pathway controls TRAF adaptor stability to regulate inflammation. *Nat Immunol* 14: 470-479.

206. Frescas, D., and M. Pagano. 2008. Deregulated proteolysis by the F-box proteins SKP2 and beta-TrCP: tipping the scales of cancer. *Nat Rev Cancer* 8: 438-449.

207. Skaar, J. R., J. K. Pagan, and M. Pagano. 2014. SCF ubiquitin ligase-targeted therapies. *Nat Rev Drug Discov* 13: 889-903.

208. Klein, C., and L. T. Vassilev. 2004. Targeting the p53-MDM2 interaction to treat cancer. *Br J Cancer* 91: 1415-1419.

209. Deshaies, R. J. 1999. SCF and Cullin/Ring H2-based ubiquitin ligases. *Annual review of cell and developmental biology* 15: 435-467.

210. Koepp, D. M., J. W. Harper, and S. J. Elledge. 1999. How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell* 97: 431-434.

211. Miyamoto, T., A. Horiuchi, H. Kashima, A. Suzuki, T. Yamada, M. Kurai, I. Konishi, and T. Shiozawa. 2010. Inverse correlation between Skp2 and p27(Kip1) in normal endometrium and endometrial carcinoma. *Gynecol Endocrinol* 26: 220-229.

212. Fukuchi, M., N. Masuda, M. Nakajima, Y. Fukai, T. Miyazaki, H. Kato, and H. Kuwano. 2004. Inverse correlation between expression levels of p27 and the ubiquitin ligase subunit Skp2 in early esophageal squamous cell carcinoma. *Anticancer Res* 24: 777-783.

213. Li, Q., M. Murphy, J. Ross, C. Sheehan, and J. A. Carlson. 2004. Skp2 and p27kip1 expression in melanocytic nevi and melanoma: an inverse relationship. *J Cutan Pathol* 31: 633-642.

214. Gao, D., H. Inuzuka, A. Tseng, R. Y. Chin, A. Toker, and W. Wei. 2009. Phosphorylation by Akt1 promotes cytoplasmic localization of Skp2 and impairs APCCdh1-mediated Skp2 destruction. *Nat Cell Biol* 11: 397-408.

215. Cardozo, T., and M. Pagano. 2004. The SCF ubiquitin ligase: insights into a molecular machine. *Nat Rev Mol Cell Biol* 5: 739-751.

216. Ganoth, D., G. Bornstein, T. K. Ko, B. Larsen, M. Tyers, M. Pagano, and A. Hershko. 2001. The cell-cycle regulatory protein Cks1 is required for SCF(Skp2)-mediated ubiquitinylation of p27. *Nat Cell Biol* 3: 321-324.

217. Spruck, C., H. Strohmaier, M. Watson, A. P. Smith, A. Ryan, T. W. Krek, and S. I. Reed. 2001. A CDK-independent function of mammalian Cks1: targeting of SCF(Skp2) to the CDK inhibitor p27Kip1. *Mol Cell* 7: 639-650.

218. Dobashi, Y., H. Tsubochi, K. Minegishi, M. Kitagawa, S. Otani, and A. Ooi. 2017. Regulation of p27 by ubiquitin ligases and its pathological significance in human lung carcinomas. *Hum Pathol*.

219. Evans, L., L. Chen, G. Milazzo, S. Gherardi, G. Perini, E. Willmore, D. R. Newell, and D. A. Tweddle. 2015. SKP2 is a direct transcriptional target of MYCN and a potential therapeutic target in neuroblastoma. *Cancer letters* 363: 37-45.

220. Schiffer, D., P. Cavalla, V. Fiano, C. Ghimenti, and R. Piva. 2002. Inverse relationship between p27/Kip.1 and the F-box protein Skp2 in human astrocytic gliomas by immunohistochemistry and Western blot. *Neurosci Lett* 328: 125-128.

221. Lu, H., X. Cao, H. Zhang, G. Sun, G. Fan, L. Chen, and S. Wang. 2014. Imbalance between MMP-2, 9 and TIMP-1 promote the invasion and metastasis of renal cell carcinoma via SKP2 signaling pathways. *Tumour Biol* 35: 9807-9813.

222. Chen, H., X. Mo, J. Yu, S. Huang, Z. Huang, and L. Gao. 2014. Interference of Skp2 effectively inhibits the development and metastasis of colon carcinoma. *Mol Med Rep* 10: 1129-1135.

223. Tian, Y. F., T. J. Chen, C. Y. Lin, L. T. Chen, L. C. Lin, C. H. Hsing, S. W. Lee, M. J. Sheu, H. H. Lee, Y. L. Shiue, H. Y. Huang, H. Y. Pan, C. F. Li, and S. H. Chen. 2013. SKP2 overexpression is associated with a poor prognosis of

rectal cancer treated with chemoradiotherapy and represents a therapeutic target with high potential. *Tumour Biol* 34: 1107-1117.

224. Zhang, B., L. H. Ji, W. Liu, G. Zhao, and Z. Y. Wu. 2013. Skp2-RNAi suppresses proliferation and migration of gallbladder carcinoma cells by enhancing p27 expression. *World J Gastroenterol* 19: 4917-4924.

225. Lv, A., Z. Li, X. Tian, X. Guan, M. Zhao, B. Dong, and C. Hao. 2013. SKP2 high expression, KIT exon 11 deletions, and gastrointestinal bleeding as predictors of poor prognosis in primary gastrointestinal stromal tumors. *PloS one* 8: e62951.

226. Liang, Y., X. Hou, Q. Cui, T. B. Kang, J. H. Fu, L. J. Zhang, R. Z. Luo, J. H. He, Y. X. Zeng, and H. X. Yang. 2012. Skp2 expression unfavorably impacts survival in resectable esophageal squamous cell carcinoma. *J Transl Med* 10: 73.

227. Kudo, Y., S. Kitajima, S. Sato, M. Miyauchi, I. Ogawa, and T. Takata. 2001. High expression of S-phase kinase-interacting protein 2, human F-box protein, correlates with poor prognosis in oral squamous cell carcinomas. *Cancer Res* 61: 7044-7047.

228. Ding, L., R. Li, X. Han, Y. Zhou, H. Zhang, Y. Cui, W. Wang, and J. Bai. 2017. Inhibition of Skp2 suppresses the proliferation and invasion of osteosarcoma cells. *Oncology reports* 38: 933-940.

229. Fu, H. C., Y. C. Yang, Y. J. Chen, H. Lin, Y. C. Ou, C. C. Chien, E. Y. Huang, H. Y. Huang, J. Lan, H. P. Chi, K. E. Huang, and H. Y. Kang. 2016.

Increased expression of SKP2 is an independent predictor of locoregional recurrence in cervical cancer via promoting DNA-damage response after irradiation. *Oncotarget* 7: 44047-44061.

230. Yang, Q., J. Huang, Q. Wu, Y. Cai, L. Zhu, X. Lu, S. Chen, C. Chen, and Z. Wang. 2014. Acquisition of epithelial-mesenchymal transition is associated with Skp2 expression in paclitaxel-resistant breast cancer cells. *Br J Cancer* 110: 1958-1967.

231. Gao, H., M. Yang, and X. Zhang. 2018. Investigating a multigene prognostic assay based on significant pathways for Luminal A breast cancer through gene expression profile analysis. . *Oncology Letters* 0.

232. Rose, A. E., G. Wang, D. Hanniford, S. Monni, T. Tu, R. L. Shapiro, R. S. Berman, A. C. Pavlick, M. Pagano, F. Darvishian, M. Mazumdar, E. Hernando, and I. Osman. 2011. Clinical relevance of SKP2 alterations in metastatic melanoma. *Pigment Cell Melanoma Res* 24: 197-206.

233. Gao, J., B. A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S. O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, E. Cerami, C. Sander, and N. Schultz. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 6: pl1.

234. Chan, C. H., C. F. Li, W. L. Yang, Y. Gao, S. W. Lee, Z. Feng, H. Y. Huang, K. K. Tsai, L. G. Flores, Y. Shao, J. D. Hazle, D. Yu, W. Wei, D. Sarbassov, M. C. Hung, K. I. Nakayama, and H. K. Lin. 2012. The Skp2-SCF E3 ligase

regulates Akt ubiquitination, glycolysis, herceptin sensitivity, and tumorigenesis. *Cell* 149: 1098-1111.

235. Lee, S. W., C. F. Li, G. Jin, Z. Cai, F. Han, C. H. Chan, W. L. Yang, B. K. Li, A. H. Rezaeian, H. Y. Li, H. Y. Huang, and H. K. Lin. 2015. Skp2-Dependent Ubiquitination and Activation of LKB1 Is Essential for Cancer Cell Survival under Energy Stress. *Mol Cell* 57: 1022-1033.

236. Bencivenga, D., I. Caldarelli, E. Stampone, F. P. Mancini, M. L. Balestrieri, F. Della Ragione, and A. Borriello. 2017. p27(Kip1) and human cancers: A reappraisal of a still enigmatic protein. *Cancer Lett* 403: 354-365.

237. Zehir, A., R. Benayed, R. H. Shah, A. Syed, S. Middha, H. R. Kim, P. Srinivasan, J. Gao, D. Chakravarty, S. M. Devlin, M. D. Hellmann, D. A. Barron, A. M. Schram, M. Hameed, S. Dogan, D. S. Ross, J. F. Hechtman, D. F. DeLair, J. Yao, D. L. Mandelker, D. T. Cheng, R. Chandramohan, A. S. Mohanty, R. N. Ptashkin, G. Jayakumaran, M. Prasad, M. H. Syed, A. B. Rema, Z. Y. Liu, K. Nafa, L. Borsu, J. Sadowska, J. Casanova, R. Bacares, I. J. Kiecka, A. Razumova, J. B. Son, L. Stewart, T. Baldi, K. A. Mullaney, H. Al-Ahmadie, E. Vakiani, A. A. Abeshouse, A. V. Penson, P. Jonsson, N. Camacho, M. T. Chang, H. H. Won, B. E. Gross, R. Kundra, Z. J. Heins, H. W. Chen, S. Phillips, H. Zhang, J. Wang, A. Ochoa, J. Wills, M. Eubank, S. B. Thomas, S. M. Gardos, D. N. Reales, J. Galle, R. Durany, R. Cambria, W. Abida, A. Cercek, D. R. Feldman, M. M. Gounder, A. A. Hakimi, J. J. Harding, G. Iyer, Y. Y. Janjigian, E. J. Jordan, C. M. Kelly, M. A. Lowery, L.

G. T. Morris, A. M. Omuro, N. Raj, P. Razavi, A. N. Shoushtari, N. Shukla, T. E. Soumerai, A. M. Varghese, R. Yaeger, J. Coleman, B. Bochner, G. J. Riely, L. B. Saltz, H. I. Scher, P. J. Sabbatini, M. E. Robson, D. S. Klimstra, B. S. Taylor, J. Baselga, N. Schultz, D. M. Hyman, M. E. Arcila, D. B. Solit, M. Ladanyi, and M. F. Berger. 2017. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 23: 703-713.

238. Kamura, T., T. Hara, M. Matsumoto, N. Ishida, F. Okumura, S. Hatakeyama, M. Yoshida, K. Nakayama, and K. I. Nakayama. 2004. Cytoplasmic ubiquitin ligase KPC regulates proteolysis of p27(Kip1) at G1 phase. *Nat Cell Biol* 6: 1229-1235.

239. Ishida, N., T. Hara, T. Kamura, M. Yoshida, K. Nakayama, and K. I. Nakayama. 2002. Phosphorylation of p27Kip1 on serine 10 is required for its binding to CRM1 and nuclear export. *J Biol Chem* 277: 14355-14358.

240. Denicourt, C., C. C. Saenz, B. Datnow, X. S. Cui, and S. F. Dowdy. 2007. Relocalized p27Kip1 tumor suppressor functions as a cytoplasmic metastatic oncogene in melanoma. *Cancer Res* 67: 9238-9243.

241. 2006. DiverSet; ChemBridge Corp.: San Diego, CA.

242. Hao, B., N. Zheng, B. A. Schulman, G. Wu, J. J. Miller, M. Pagano, and N. P. Pavletich. 2005. Structural basis of the Cks1-dependent recognition of p27(Kip1) by the SCF(Skp2) ubiquitin ligase. *Mol Cell* 20: 9-19.

243. Schulman, B. A., A. C. Carrano, P. D. Jeffrey, Z. Bowen, E. R. Kinnucan, M. S. Finnin, S. J. Elledge, J. W. Harper, M. Pagano, and N. P. Pavletich. 2000. Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex. *Nature* 408: 381-386.

244. Morrow, J. K., and S. Zhang. 2012. Computational prediction of protein hot spot residues. *Curr Pharm Des* 18: 1255-1265.

245. Zheng, N., B. A. Schulman, L. Song, J. J. Miller, P. D. Jeffrey, P. Wang, C. Chu, D. M. Koepp, S. J. Elledge, M. Pagano, R. C. Conaway, J. W. Conaway, J. W. Harper, and N. P. Pavletich. 2002. Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature* 416: 703-709.

246. Sussman, J. L., D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter, and E. E. Abola. 1998. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallogr D Biol Crystallogr* 54: 1078-1084.

247. 2009. The PyMOL Molecular Graphics System, Version 1.2r1, Schr”dinger, LLC.

248. 2009. Molecular Operating Environment (MOE), developed and distributed by Chemical Computing Group http://www.chemcomp.com/software-moe2009.htm

249. Zhang, S., and L. Du-Cuny. 2009. Development and evaluation of a new statistical model for structure-based high-throughput virtual screening. *Int J Bioinform.Res Appl.* 5: 269-279.

250. Zhang, S., K. Kumar, X. Jiang, A. Wallqvist, and J. Reifman. 2008. DOVIS: an implementation for high-throughput virtual screening using AutoDock. *BMC.Bioinformatics.* 9: 126.

251. Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. 2003. Improved protein-ligand docking using GOLD. *Proteins* 52: 609-623.

252. Yang, W. L., J. Wang, C. H. Chan, S. W. Lee, A. D. Campos, B. Lamothe, L. Hur, B. C. Grabiner, X. Lin, B. G. Darnay, and H. K. Lin. 2009. The E3 ligase TRAF6 regulates Akt ubiquitination and activation. *Science* 325: 1134-1138.

253. Li, T., N. P. Pavletich, B. A. Schulman, and N. Zheng. 2005. High-level expression and purification of recombinant SCF ubiquitin ligases. *Methods Enzymol* 398: 125-142.

254. Wang, L., W. Zhang, Q. Gao, and C. Xiong. 2014. Prediction of hot spots in protein interfaces using extreme learning machines with the information of spatial neighbour residues. *IET Syst Biol* 8: 184-190.

255. Hui-fang, L., S. Qing, Z. Jian, and F. Wei. 2010. Evaluation of various inverse docking schemes in multiple targets identification. *J Mol Graph Model* 29: 326-330.

256. Plewczynski, D., M. Lazniewski, R. Augustyniak, and K. Ginalski. 2011. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J Comput Chem* 32: 742-755.

257. Li, X., Y. Li, T. Cheng, Z. Liu, and R. Wang. 2010. Evaluation of the performance of four molecular docking programs on a diverse set of protein-ligand complexes. *J Comput Chem* 31: 2109-2125.

258. Suzuki, H., T. Chiba, T. Suzuki, T. Fujita, T. Ikenoue, M. Omata, K. Furuichi, H. Shikama, and K. Tanaka. 2000. Homodimer of two F-box proteins betaTrCP1 or betaTrCP2 binds to IkappaBalpha for signal-dependent ubiquitination. *J Biol Chem* 275: 2877-2884.

259. Wang, L., X. Ye, Y. Liu, W. Wei, and Z. Wang. 2014. Aberrant regulation of FBW7 in cancer. *Oncotarget* 5: 2000-2015.

260. Sigrist, C. J., E. de Castro, L. Cerutti, B. A. Cuche, N. Hulo, A. Bridge, L. Bougueleret, and I. Xenarios. 2013. New and continuing developments at PROSITE. *Nucleic Acids Res* 41: D344-347.

261. Huang, H. C., C. L. Lin, and J. K. Lin. 2011. 1,2,3,4,6-penta-O-galloyl-beta-D-glucose, quercetin, curcumin and lycopene induce cell-cycle arrest in MDA-MB-231 and BT474 cells through downregulation of Skp2 protein. *J Agric Food Chem* 59: 6765-6775.

262. Chen, X. M., Y. Bai, Y. J. Zhong, X. L. Xie, H. W. Long, Y. Y. Yang, S. G. Wu, Q. Jia, and X. H. Wang. 2013. Wogonin has multiple anti-cancer effects by regulating c-Myc/SKP2/Fbw7alpha and HDAC1/HDAC2 pathways and inducing apoptosis in human lung adenocarcinoma cell line A549. *PLoS One* 8: e79201.

263. Zhao, K., L. Wei, H. Hui, Q. Dai, Q. D. You, Q. L. Guo, and N. Lu. 2014. Wogonin suppresses melanoma cell B16-F10 invasion and migration by inhibiting Ras-medicated pathways. *PLoS One* 9: e106458.

264. Hong, M., H. Cheng, L. Song, W. Wang, Q. Wang, D. Xu, and W. Xing. 2018. Wogonin Suppresses the Activity of Matrix Metalloproteinase-9 and Inhibits Migration and Invasion in Human Hepatocellular Carcinoma. *Molecules* 23.

265. Meng, S., B. Wu, R. Singh, T. Yin, J. K. Morrow, S. Zhang, and M. Hu. 2012. SULT1A3-mediated regiospecific 7-O-sulfation of flavonoids in Caco-2 cells can be explained by the relevant molecular docking studies. *Mol Pharm* 9: 862-873.

266. Chen, J., H. Lin, and M. Hu. 2005. Absorption and metabolism of genistein and its five isoflavone analogs in the human intestinal Caco-2 model. *Cancer Chemother Pharmacol* 55: 159-169.

267. Vervoorts, J., and B. Luscher. 2008. Post-translational regulation of the tumor suppressor p27(KIP1). *Cell Mol Life Sci* 65: 3255-3264.

268. Cavasotto, C. N., and S. S. Phatak. 2009. Homology modeling in drug discovery: current trends and applications. *Drug Discov Today* 14: 676-683.

269. Daga, P. R., R. Y. Patel, and R. J. Doerksen. 2010. Template-based protein modeling: recent methodological advances. *Curr Top Med Chem* 10: 84-94.

270. Malmstrom, L., and D. R. Goodlett. 2010. Protein structure modeling. *Methods Mol Biol* 673: 63-72.

271. Grant, M. A. 2009. Protein structure prediction in structure-based ligand design and virtual screening. *Comb Chem High Throughput Screen* 12: 940-960.

272. Liu, X., D. Shi, S. Zhou, H. Liu, H. Liu, and X. Yao. 2018. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov* 13: 23-37.

273. Hao, B., S. Oehlmann, M. E. Sowa, J. W. Harper, and N. P. Pavletich. 2007. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol Cell* 26: 131-143.

274. Sali, A., and T. L. Blundell. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815.

275. Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. 2005. The Amber biomolecular simulation programs. *J Comput Chem* 26: 1668-1688.

276. Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 11: 3696-3713.

277. Mahoney, M. W., and W. L. Jorgensen. 2000. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J Chem Phys* 112: 8910-8922.

278. Roe, D. R., and T. E. Cheatham, 3rd. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 9: 3084-3095.

279. Kim, S., P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang, and S. H. Bryant. 2016. PubChem Substance and Compound databases. *Nucleic Acids Res* 44: D1202-1213.

280. Huang, H., K. M. Regan, F. Wang, D. Wang, D. I. Smith, J. M. van Deursen, and D. J. Tindall. 2005. Skp2 inhibits FOXO1 in tumor suppression through ubiquitin-mediated degradation. *Proceedings of the National Academy of Sciences of the United States of America* 102: 1649-1654.

281. Hnit, S. S., C. Xie, M. Yao, J. Holst, A. Bensoussan, P. De Souza, Z. Li, and Q. Dong. 2015. p27(Kip1) signaling: Transcriptional and post-translational regulation. *Int J Biochem Cell Biol* 68: 9-14.

282. Rahman, K., P. Zhao, M. Mandalasi, H. van der Wel, L. Wells, I. J. Blader, and C. M. West. 2016. The E3 Ubiquitin Ligase Adaptor Protein Skp1 Is Glycosylated by an Evolutionarily Conserved Pathway That Regulates Protist Growth and Development. *J Biol Chem* 291: 4268-4280.

283. Sheikh, M. O., Y. Xu, H. van der Wel, P. Walden, S. D. Hartson, and C. M. West. 2015. Glycosylation of Skp1 promotes formation of Skp1-cullin-1-F-box protein complexes in dictyostelium. *Mol Cell Proteomics* 14: 66-80.

284. Wada, H., E. T. Yeh, and T. Kamitani. 1999. Identification of NEDD8-conjugation site in human cullin-2. *Biochemical and biophysical research communications* 257: 100-105.

285. Morimoto, M., T. Nishida, R. Honda, and H. Yasuda. 2000. Modification of cullin-1 by ubiquitin-like protein Nedd8 enhances the activity of SCF(skp2) toward p27(kip1). *Biochem Biophys Res Commun* 270: 1093-1096.

286. Enchev, R. I., B. A. Schulman, and M. Peter. 2015. Protein neddylation: beyond cullin-RING ligases. *Nat Rev Mol Cell Biol* 16: 30-44.

287. Lydeard, J. R., B. A. Schulman, and J. W. Harper. 2013. Building and remodelling Cullin-RING E3 ubiquitin ligases. *EMBO Rep* 14: 1050-1061.

288. Duda, D. M., D. C. Scott, M. F. Calabrese, E. S. Zimmerman, N. Zheng, and B. A. Schulman. 2011. Structural regulation of cullin-RING ubiquitin ligase complexes. *Curr Opin Struct Biol* 21: 257-264.

289. Inuzuka, H., D. Gao, L. W. Finley, W. Yang, L. Wan, H. Fukushima, Y. R. Chin, B. Zhai, S. Shaik, A. W. Lau, Z. Wang, S. P. Gygi, K. Nakayama, J. Teruya-Feldstein, A. Toker, M. C. Haigis, P. P. Pandolfi, and W. Wei. 2012. Acetylation-dependent regulation of Skp2 function. *Cell* 150: 179-193.

290. Obata, T., M. B. Yaffe, G. G. Leparc, E. T. Piro, H. Maegawa, A. Kashiwagi, R. Kikkawa, and L. C. Cantley. 2000. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *The Journal of biological chemistry* 275: 36108-36115.

291. Lin, H. K., G. Wang, Z. Chen, J. Teruya-Feldstein, Y. Liu, C. H. Chan, W. L. Yang, H. Erdjument-Bromage, K. I. Nakayama, S. Nimer, P. Tempst, and P. P. Pandolfi. 2009. Phosphorylation-dependent regulation of cytosolic localization and oncogenic function of Skp2 by Akt/PKB. *Nat Cell Biol* 11: 420-432.

292. Rodier, G., P. Coulombe, P. L. Tanguay, C. Boutonnet, and S. Meloche. 2008. Phosphorylation of Skp2 regulated by CDK2 and Cdc14B protects it from degradation by APC(Cdh1) in G1 phase. *The EMBO journal* 27: 679-691.

293. Bashir, T., J. K. Pagan, L. Busino, and M. Pagano. 2010. Phosphorylation of Ser72 is dispensable for Skp2 assembly into an active SCF ubiquitin ligase and its subcellular localization. *Cell cycle* 9: 971-974.

294. Boutonnet, C., P. L. Tanguay, C. Julien, G. Rodier, P. Coulombe, and S. Meloche. 2010. Phosphorylation of Ser72 does not regulate the ubiquitin ligase activity and subcellular localization of Skp2. *Cell cycle* 9: 975-979.

295. Zhou, H., S. Di Palma, C. Preisinger, M. Peng, A. N. Polat, A. J. Heck, and S. Mohammed. 2013. Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res* 12: 260-271.

296. Gao, D., H. Inuzuka, A. Tseng, and W. Wei. 2009. Akt finds its new path to regulate cell cycle through modulating Skp2 activity and its destruction by APC/Cdh1. *Cell Div* 4: 11.

297. Geng, Q., J. Liu, Z. Gong, S. Chen, S. Chen, X. Li, Y. Lu, X. Zhu, H. K. Lin, and D. Xu. 2017. Phosphorylation by mTORC1 stablizes Skp2 and regulates its oncogenic function in gastric cancer. *Mol Cancer* 16: 83.

298. Zachariae, W., and K. Nasmyth. 1999. Whose end is destruction: cell division and the anaphase-promoting complex. *Genes & development* 13: 2039-2058.

299. Wei, W., N. G. Ayad, Y. Wan, G. J. Zhang, M. W. Kirschner, and W. G. Kaelin, Jr. 2004. Degradation of the SCF component Skp2 in cell-cycle phase G1 by the anaphase-promoting complex. *Nature* 428: 194-198.

300. Najmanovich, R., J. Kuttner, V. Sobolev, and M. Edelman. 2000. Side-chain flexibility in proteins upon ligand binding. *Proteins* 39: 261-268.

301. Moras, D., and H. Gronemeyer. 1998. The nuclear receptor ligand-binding domain: structure and function. *Current opinion in cell biology* 10: 384-391.

302. Skwarczynska, M., and C. Ottmann. 2015. Protein-protein interactions as drug targets. *Future Med Chem* 7: 2195-2219.

303. Makley, L. N., and J. E. Gestwicki. 2013. Expanding the number of 'druggable' targets: non-enzymes and protein-protein interactions. *Chem Biol Drug Des* 81: 22-32.

304. Guo, W., J. A. Wisniewski, and H. Ji. 2014. Hot spot-based design of small-molecule inhibitors for protein-protein interactions. *Bioorg Med Chem Lett* 24: 2546-2554.

305. Ge, H., Y. Wang, C. Li, N. Chen, Y. Xie, M. Xu, Y. He, X. Gu, R. Wu, Q. Gu, L. Zeng, and J. Xu. 2013. Molecular dynamics-based virtual screening:

accelerating the drug discovery process by high-performance computing. *J Chem Inf Model* 53: 2757-2764.

306. Kutzner, C., S. Pall, M. Fechner, A. Esztermann, B. L. de Groot, and H. Grubmuller. 2015. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J Comput Chem* 36: 1990-2008.

307. Purawat, S., P. U. Ieong, R. D. Malmstrom, G. J. Chan, A. K. Yeung, R. C. Walker, I. Altintas, and R. E. Amaro. 2017. A Kepler Workflow Tool for Reproducible AMBER GPU Molecular Dynamics. *Biophys J* 112: 2469-2474.

308. Martinez-Rosell, G., T. Giorgino, M. J. Harvey, and G. de Fabritiis. 2017. Drug Discovery and Molecular Dynamics: Methods, Applications and Perspective Beyond the Second Timescale. *Curr Top Med Chem* 17: 2617-2625.

309. Rotin, D., and S. Kumar. 2009. Physiological functions of the HECT family of ubiquitin ligases. *Nat Rev Mol Cell Biol* 10: 398-409.

310. Le, X. F., W. Mao, G. He, F. X. Claret, W. Xia, A. A. Ahmed, M. C. Hung, Z. H. Siddik, and R. C. Bast, Jr. 2011. The role of p27(Kip1) in dasatinib-enhanced paclitaxel cytotoxicity in human ovarian cancer cells. *J Natl Cancer Inst* 103: 1403-1422.

311. Le, X. F., F. Pruefer, and R. C. Bast, Jr. 2005. HER2-targeting antibodies modulate the cyclin-dependent kinase inhibitor p27Kip1 via multiple signaling pathways. *Cell Cycle* 4: 87-95.

312. Le, X. F., F. X. Claret, A. Lammayot, L. Tian, D. Deshpande, R. LaPushin, A. M. Tari, and R. C. Bast, Jr. 2003. The role of cyclin-dependent kinase inhibitor p27Kip1 in anti-HER2 antibody-induced G1 cell cycle arrest and tumor growth inhibition. *J Biol Chem* 278: 23441-23450.

313. He, G., J. Kuang, Z. Huang, J. Koomen, R. Kobayashi, A. R. Khokhar, and Z. H. Siddik. 2006. Upregulation of p27 and its inhibition of CDK2/cyclin E activity following DNA damage by a novel platinum agent are dependent on the expression of p21. *Br J Cancer* 95: 1514-1524.

314. Shirane, M., Y. Harumiya, N. Ishida, A. Hirai, C. Miyamoto, S. Hatakeyama, K. Nakayama, and M. Kitagawa. 1999. Down-regulation of p27(Kip1) by two mechanisms, ubiquitin-mediated degradation and proteolytic processing. *J Biol Chem* 274: 13886-13893.

315. Shoemaker, R. H. 2006. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 6: 813-823.

316. Hughes, T. R., M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, and S. H. Friend. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102: 109-126.

317. Gardner, T. S., D. di Bernardo, D. Lorenz, and J. J. Collins. 2003. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 301: 102-105.

318. Lamb, J., E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J. P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub. 2006. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313: 1929-1935.

319. Hieronymus, H., J. Lamb, K. N. Ross, X. P. Peng, C. Clement, A. Rodina, M. Nieto, J. Du, K. Stegmaier, S. M. Raj, K. N. Maloney, J. Clardy, W. C. Hahn, G. Chiosis, and T. R. Golub. 2006. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* 10: 321-330.

320. Jiang, H., J. R. Pritchard, R. T. Williams, D. A. Lauffenburger, and M. T. Hemann. 2011. A mammalian functional-genetic approach to characterizing cancer therapeutics. *Nat Chem Biol* 7: 92-100.

Vita

John Kenneth Morrow was born in Woodland, California, the son of Philip Ross Morrow and Jeanne Morrow. After completing his Bachelor of Science degree at University of California at Berkeley in 2007 in Molecular Toxicology, he obtained his Master's degree from University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences at Houston in Experimental Therapeutics in 2012. In September of 2012, he entered the Experimental Therapeutics Ph.D. program in the University Of Texas MD Anderson Cancer Center UTHealth Graduate School Of Biomedical Sciences at Houston.